

Diplomatura d'Estadística

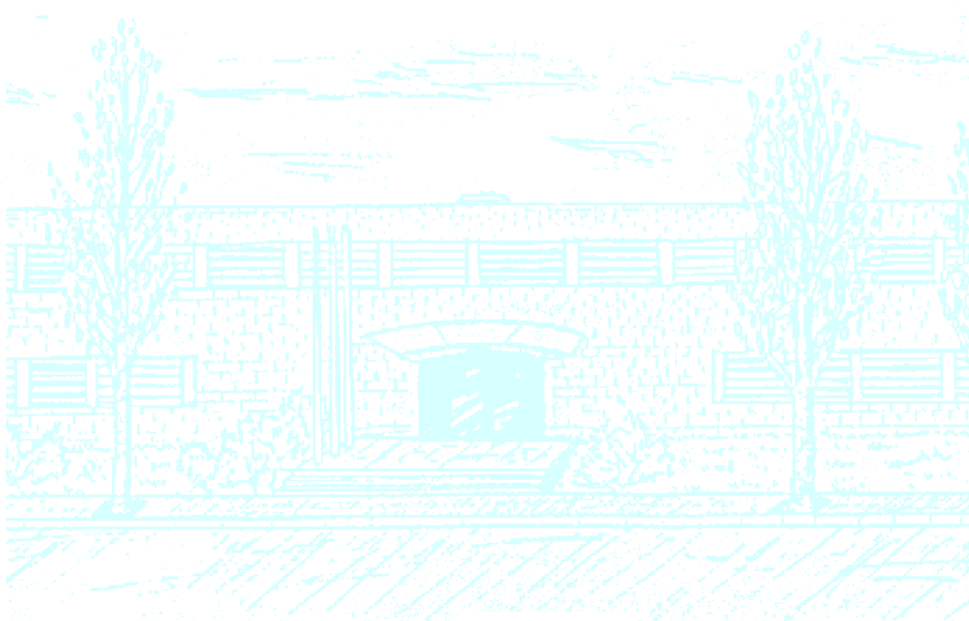
Títol: Evolució de l'IMC en la població espanyola entre 1995 i 2006

Autor: Anna Fernández Blanchart

Directora: Lourdes Rodero de Lamo

Departament: Departament d'Estadística i Investigació Operativa

Convocatòria: Juny del 2011



Facultat de Matemàtiques
i Estadística

UNIVERSITAT POLITÈCNICA DE CATALUNYA

Facultat de Matemàtiques i Estadística
Universitat Politècnica de Catalunya

Projecte final de carrera

Evolució de l'IMC en la població espanyola entre 1995 i 2006

Anna Fernández Blanchart

Directora: Lourdes Rodero de Lamo

Departament d'Estadística i Investigació Operativa

És millor resoldre el problema correcte aproximadament
que resoldre exactament el problema equivoccat.

(J. W. Tukey)

1. ÍNDEX

| | |
|--|----|
| 2. INTRODUCCIÓ | 2 |
| 3. DESCRIPCIÓ DE LES BASES DE DADES..... | 4 |
| 3.1. ENQUESTA NACIONAL DE SALUT..... | 4 |
| 3.1.1. Descripció de l'enquesta | 4 |
| 3.1.2. Selecció de la mostra..... | 6 |
| 3.1.3. Descripció de la base de dades | 9 |
| 3.2. DESCRIPCIÓ DE LA VARIABLE D'INTERÈS. L'ÍNDEX DE MASSA CORPORAL..... | 10 |
| 4. ANÀLISI DESCRIPTIVA | 15 |
| 4.1. ANY 1995..... | 16 |
| 4.1.1. Depuració de la base de dades | 17 |
| 4.1.2. Descriptiva univariant..... | 20 |
| 4.1.3. Descriptiva bivariant..... | 32 |
| 4.2. ANY 2006..... | 37 |
| 4.2.1. Depuració de la base de dades | 37 |
| 4.2.2. Descriptiva univariant..... | 40 |
| 4.2.3. Descriptiva bivariant..... | 50 |
| 5. COMPARATIVA DE LES DISTRIBUCIONS DE L'IMC DEL 1995 I 2006 .. | 55 |
| 5.1. AJUST DE DISTRIBUCIONS..... | 55 |
| 5.2. FUNCIONS DE DENSITAT | 57 |
| 5.3. TUKEY MEAN-DIFFERENCE PLOT..... | 61 |

| | |
|--|-----|
| 5.4. COMPARATIVA ENTRE ESPANYA I EE.UU. | 73 |
| 6. CORBES CORRESPONENTS ALS QUARTILS DE L'IMC | 75 |
| 7. CONCLUSIONS | 81 |
| 8. BIBLIOGRAFIA..... | 83 |
| 9. ANNEX | 84 |
| 9.1. NOTÍCIA SOBRE LES ENSE..... | 84 |
| 9.2. SCRIPT DE L'R..... | 85 |
| 9.3. QUESTIONARI CORRESPONENT ALS ADULTS DE L'ENSE DE L'ANY 1995..... | 98 |
| 9.4. QUESTIONARI CORRESPONENT ALS ADULTS DE L'ENSE DE L'ANY 2006..... | 99 |
| 9.5. ARTÍCLE NORD-AMERICÀ: "Changes in the distribution og body mass index of adults and children in the US population" | 100 |

2. INTRODUCCIÓ

El problema del sobrepès preocupa i envaeix la nostra societat. Sembla que l'obesitat, cada cop, afecta a un major nombre de persones i que el pes de la població tendeix a augmentar. Aquest fet, unit a que recentment s'han publicat per primera vegada les microdades de les enquestes nacionals de salut espanyoles (ENSE), ha fet que ens semblés força interessant i oportú el tractament d'aquest tema mitjançant un projecte final de carrera.

Fins ara només s'havien mostrat els resultats dels estudis elaborats per l'INE a partir de les dades de les ENSE; però mai s'havien fet públiques les bases de dades amb els registres individuals (tal i com demostra la notícia adjuntada en l'annex 9.1.). Això feia impossible l'elaboració d'altres estudis paral·lels, com aquest. A més, cal tenir en compte que és molt difícil aconseguir una base de dades d'aquestes característiques, així que ens va semblar una gran oportunitat.

Aquest projecte té com a principal objectiu investigar els possibles canvis entre les distribucions de l'índex de massa corporal (IMC) espanyol del 1995 i el del 2006, amb la intenció de veure si tendim a estar més grassos. A més de comparar les distribucions globals dels l'IMC del 1995 i el 2006, també ens proposem fer-ho estratificant per edats i per sexe amb la intenció de detectar i analitzar les diferències. Hem escollit les dades amb 10 anys de diferència perquè creiem que així ja podrem detectar diferències en cas de que n'hi hagi.

Un objectiu secundari que ens hem marcat és construir una eina gràfica que representi els percentils poblacionals dels adults permetent d'una manera fàcil ubicar un individu dins la població. Amb això volem dir que es pugui aproximar a quin percentil es troba respecte el seu IMC i en quina categoria (infrapès, normopès, preobesitat o obesitat). La idea, doncs, és fer un símil a les corbes que utilitzen els pediatres per avaluar el creixement dels seus pacients (es presenten en el capítol 3).

Prèviament a l'inici del projecte en sí, es va dur a terme una cerca bibliogràfica i d'estat de l'art, durant la qual es va trobar un article nord-americà (annex 9.5.) amb objectius similars als nostres que hem utilitzat per comparar resultats, entre d'altres. Al treballar-lo a fons vam copsar que proposaven un manera d'abordar el problema molt interessant i vam decidir inspirar-nos en ells en certs aspectes del nostre anàlisi, treballant de manera similar a com ho fan ells. A més, això ens ha permès fer comparatives molt riques ja que tenim un estudi amb unes característiques similars al nostre, d'un país acusa una problemàtica important pel que fa a l'obesitat; fent-nos més fàcil veure quin és l'estat de la nostra població.

Previ a tot això s'ha fet un anàlisi bivariant i univariant de les variables amb les que treballarem, com sempre és recomanat per veure que succeeix i detectar i eliminar anomalies si és necessari. Aquesta anàlisi i tot el que es presenta en aquest projecte ha estat elaborat mitjançant R; programa d'accés lliure, amb una infinitat de possibilitats i que permet el tracte amb bases de dades tan grans com les nostres. En l'annex 9.2. s'adjunta l'script d'R emprat per l'elaboració de tots els anàlisis.

3. DESCRIPCIÓ DE LES BASES DE DADES

3.1. ENQUESTA NACIONAL DE SALUT

L'enquesta nacional de salut espanyola (ENSE) és una enquesta realitzada pel "Ministerio de Sanidad, Política Social e Igualdad" periòdicament des de l'any 1987. Es du a terme mitjançant entrevistes, i el seu objectiu principal és captar la informació necessària sobre la salut de la població per poder planificar i avaluar les actuacions sanitàries, procurant per la millora de l'atenció dels usuaris dels serveis sanitaris.

En els inicis d'aquesta enquesta (les primeres 5 entregues entre 1987 i 2001), l'organisme encarregat de realitzar el treball de camp era el "Centro de Investigaciones Sociológicas" (CIS). A partir de l'entrega del 2003, li va prendre el relleu l'"Instituto Nacional de Estadística" (INE), que és qui se n'ha encarregat d'aleshores ençà.

La darrera enquesta publicada en la seva totalitat, que es va fer l'any 2006, fou revisada i modificada, ampliant el nombre de preguntes per tal d'adaptar-se a l'entorn europeu. D'aquesta manera, en l'actualitat l'enquesta nacional de salut recull molta més informació i s'ha equiparat a la resta de països de la UE i a la que es fa a nivell europeu; convertint-se així en la mateixa enquesta que la europea però a nivell nacional.

3.1.1. Descripció de l'enquesta

L'enquesta nacional de salut Espanyola (ENSE) és una operació estadística realitzada conjuntament pel *Ministerio de Sanidad y Consumo* i l'*Instituto Nacional de Estadística* dins del *Plan Estadístico Nacional* amb l'objectiu de recollir informació sobre la salut des del punt de vista dels ciutadans i poder així

desenvolupar unes millors estratègies sanitàries. S'elabora aquesta enquesta amb una periodicitat de dos anys.

L'enquesta en sí consta de tres parts organitzades en dues fases: en la primera fase es passa un qüestionari a una llar en general; i en la segona s'enquesta a una persona adulta (més de 16 anys) i un menor (de 0 a 15 anys) d'aquella mateixa llar.

En aquest treball ens centrarem en les variables que s'extreuen del qüestionari dels adults, ja que abordem objectius centrats en aquest sector de la població. Aquesta part de l'enquesta consta de més de 140 preguntes organitzades en diferents blocs segons el determinant de salut del qual tracten, des de la revisió del 2003. Els diferents determinants de salut que s'investiguen són: treball reproductiu, estat de salut i malalties cròniques, accidents, restricció de l'activitat, consum de medicaments, salut mental, estrés laboral, accés i utilització dels serveis sanitaris, hàbits de vida, pràctiques preventives, agressions, discriminació, característiques físiques i sensorials, limitacions per a realitzar activitats de la vida quotidiana, suport afectiu i personal, i funció familiar. Pel que fa a les enquestes anteriors a la revisió, cal tenir en compte que no recullen una informació tant detallada; però tot i això aporten la informació suficient per abordar una comparativa com la que ens interessa en el nostre cas.

Tant en els anys anteriors a la revisió com ara, el qüestionari dels adults s'ha de contestar en totes les llars seleccionades sense excepció, passant una entrevista personal a una única persona. Només en casos molt excepcionals els entrevistadors poden completar l'entrevista per telèfon.

Es poden trobar les preguntes de les ENSE d'on s'han extret les dades per elaborar els anàlisis i els links on es poden trobar completes, en l'annex 9.3. i 9.4.. En elles també es pot observar el canvi després de la revisió del 2003, ja que la primera és del 1995 i la segona del 2006.

3.1.2. Selecció de la mostra

La població d'estudi és el conjunt de persones que resideixen en l'estat espanyol, per tant es mostreja sobre tot el territori nacional censat. El període de recollida d'informació és d'un any de durada i l'estratègia de mostreig emprada és la mateixa tant abans com després de la revisió esmentada.

El tipus de mostreig que s'utilitza és un mostreig multietàpic estratificat que consta de tres etapes (representat esquemàticament en la figura 3.1.). Les unitats de la primera etapa són les seccions censals. Les de la segona etapa són els habitatges familiars i dins de cada habitatge és seleccionat un adult i un menor (en el cas de que n'hi hagin) perquè sigui possible omplir les tres enquestes de salut nacional: la del habitatge, la dels adults i la dels nens.

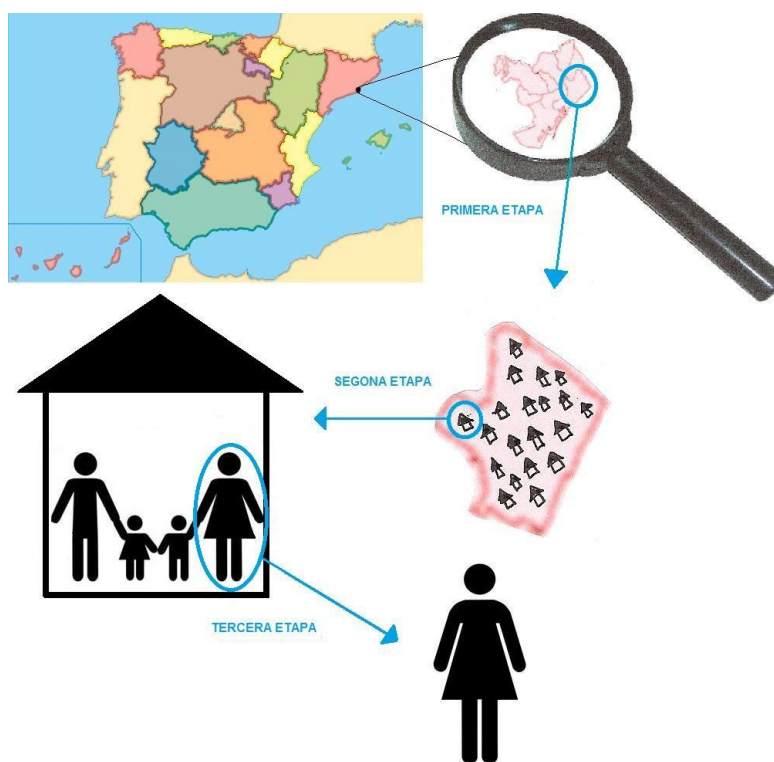


Figura 3.1. esquema gràfic del mostreig de l'ENSE.

En la primera etapa les seccions censals s'agrupen en estrats segons la mida del municipi al qual pertany cadascuna d'elles. Els estrats considerats són els següents:

Estrat 0: Municipis amb més de 500.000 habitants.

Estrat 1: Municipis capitals de província (excepte els anteriors).

Estrat 2: Municipis amb més de 100.000 habitants (excepte els anteriors).

Estrat 3: Municipis de 50.000 a 100.000 habitants (excepte els anteriors).

Estrat 4: Municipis de 20.000 a 50.000 habitants (excepte els anteriors).

Estrat 5: Municipis de 10.000 a 20.000 habitants.

Estrat 6: Municipis amb menys de 10.000 habitants.

La mostra del 2006, per exemple, consta d'uns 31.000 habitatges distribuïts en 2.236 seccions censals, ja que aquesta és la mida que permet cobrir el grau de fiabilitat de les estimacions. Com que l'estudi elaborat per l'INE pretén poder facilitar dades per cada comunitat autònoma, es dissenya una mostra independent que representa a cadascuna, condicionant així la selecció de les seccions censals.

La mostra de seccions censals és distribuïda entre les diferents comunitats autònomes assignant una part uniformement i l'altra proporcional a la mida de la comunitat. La distribució de la mostra per comunitats autònomes és la presentada en la taula 3.1.

| Comunitat Autònoma | Nombre de seccions censals |
|----------------------|----------------------------|
| Andalusia | 200 |
| Aragó | 200 |
| Astúries | 68 |
| Balears | 132 |
| Canàries | 84 |
| Cantàbria | 124 |
| Castella i Lleó | 100 |
| Castella la Manxa | 84 |
| Catalunya | 188 |
| Comunitat Valenciana | 136 |
| Extremadura | 68 |
| Galícia | 252 |
| Madrid | 164 |
| Múrcia | 140 |
| Navarra | 116 |
| País Basc | 88 |
| Rioja | 52 |
| Ceuta i Melilla | 40 |
| TOTAL | 2.236 |

Taula 3.1. Distribució de la mostra de la primera etapa (seccions censals) per comunitats autònomes.

Al mateix temps, es seleccionen les seccions censals dins de cada estrat amb una probabilitat proporcional a la seva grandària. La selecció dels habitatges dins de cada secció és equiprobable i s'elabora mitjançant un mostreig sistemàtic amb inici aleatori. Aquest procediment ens porta a tenir mostres autoponderades en cadascun dels estrats.

Pel que fa a la selecció de les persones dins dels habitatges, s'utilitza un procediment aleatori on tots els individus tenen la mateixa probabilitat de respondre l'enquesta que els pertoca.

3.1.3. Descripció de la base de dades

La base de dades amb la que es treballarà en aquest projecte és concretament la que correspon al qüestionari individual dels adults (veure annexos 9.3. i 9.4.). Utilitzarem les bases de dades de l'any 1995 i del 2006 per tal de dur a terme els objectius comparatius anteriorment exposats; que s'han escollit amb uns 10 anys de separació perquè creiem que seran prous perquè es puguin veure diferències (i per símil amb l'article de l'annex 9.5.). A cada persona li correspon un registre, i en el 1995 n'hi ha un total de 6.396. Aquesta serà doncs la longitud de la base de dades del 1995 en quan a nombre d'individus. Pel que fa al 2006 tenim un total de 29.478 individus.

En quant al nombre de variables, la bases de dades corresponent a l'any 2006 en té un total de 859. És amb aquest gran nombre de variables que es plasma tota la informació obtinguda mitjançant l'enquesta individual dels adults, amb l'objectiu d'elaborar una posterior anàlisi estadística. En el 1995 hi ha un nombre notablement menor de variables, ja que aquesta base de dades és prèvia a la modificació en l'ENSE. En aquest cas tenim un total de 64 variables.

De tot aquest conjunt de variables ens centrarem bàsicament en el pes i l'alçada, variables que conformen l'Índex de massa corporal. Aquest índex serà el centre de tot el nostre estudi, ja que estem interessats en informació relativa a la relació entre l'alçada i el pes. Altres variables que emprarem són les habituals per realitzar comparacions, com l'edat i el sexe.

3.2. DESCRIPCIÓ DE LA VARIABLE D'INTERÈS. **L'ÍNDEX DE MASSA CORPORAL**

L'índex de massa corporal (IMC) és una mesura que associa el pes amb l'alçada, permetent així donar-nos una idea de l'adipositat d'un individu determinat. Aquesta mesura fou ideada per l'estadístic Lambert Adolphe Jaques Quételet entre el 1830 i el 1850, i és per aquesta raó que també se la coneix com a índex de Quételet. Si bé fou aleshores quan es va idear, no va ser fins al 1972 quan va adquirir renom arrel d'un article escrit pel científic americà Ancel Keys.

Degut a les diferents unitats de mesura, hi ha diferents fórmules per calcular l'IMC (també conegut com BMI per les seves sigles en anglès). Donada la naturalesa de la nostra base de dades i seguint les unitats del sistema internacional (SI), calcularem l'IMC així:

$$IMC = \frac{pes\ (kg)}{(alçada\ (m))^2}, \text{ on les indicacions mostrades entre parèntesis són les unitats de mesura.}$$

Per passar l'IMC a unes altres unitats de mesura, es pot multiplicar el resultat de la fórmula anterior per les constants que es presenten a continuació (a la taula 3.2.).

| Unitats del pes | Unitats de l'alçada | Constant |
|-----------------|---------------------|----------|
| Lliure (lb) | Polzades (in) | 703 |
| Lliure (lb) | Peus (ft) | 4,88 |
| Stones (st) | Polzades (in) | 9840 |

Taula 3.2. Equivalències pel càlcul de l'IMC.

L'IMC es pot categoritzar agrupant els individus independentment del seu sexe i la seva edat en els següents grups:

| CLASSIFICACIÓ | IMC (Kg/m ²) |
|--------------------|--------------------------|
| Infrapès | < 18,50 |
| Primesa severa | <16,00 |
| Primesa moderada | 16,00 - 16,99 |
| Primesa acceptable | 17,00 – 18,49 |
| Normopès | 18,5 – 24,99 |
| Sobrepès | 25,00 – 29,99 |
| Obesitat | ≥ 30,00 |
| Obesitat tipus I | 30,00 – 34,99 |
| Obesitat tipus II | 35,00 – 39,99 |
| Obesitat tipus III | ≥ 40,00 |

Taula 3.3. Classificació del IMC europea estàndard.

En la figura 3.2 es presenta una eina gràfica que permet relacionar de manera visual, sense necessitat de càlculs, l'alçada i el pes amb la categoria de l'IMC. De fet d'alguna manera és la representació de la taula 3.3., pintant les zones corresponents a cada categoria de l'IMC segons l'alçada i el pes per evitar la necessitat de fer càlculs. És clar, però, que d'aquesta manera només es pot conèixer la categoria i no el valor numèric exacte de l'IMC.

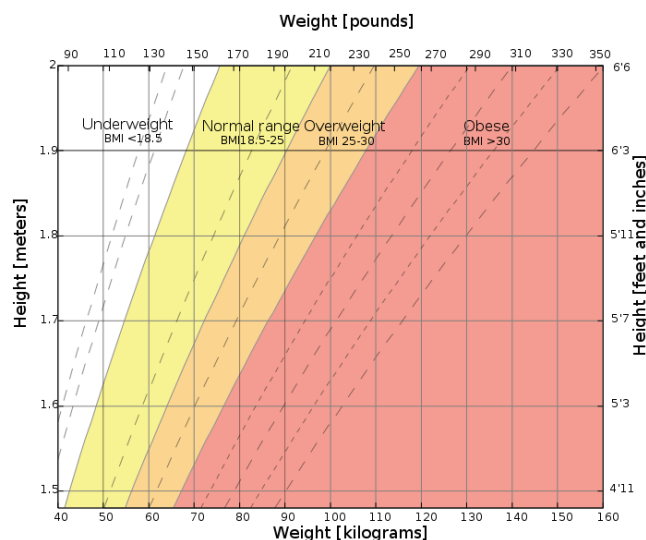


Figura 3.2. Representació gràfica de la categorització del IMC. Font: wikipedia.org

La categorització de l'IMC varia segons el lloc i el moment; ja que la constitució de persones depenen del territori, raça i de l'època (no es tenia la mateixa constitució al segle XIX que ara). La categorització presentada és l'estàndard europea de l'actualitat.

Així, s'ha determinat que cal canviar els líndars de l'IMC segons la població estudiada si no es vol estar cometent alguns errors. Un exemple molt clar és el dels orientals; pels quals cal usar una altra categorització d'aquest índex donat que són persones de complexió més petita. Així, si no es vol estar cometent errors en la classificació d'orientals, cal emprar una categorització adient a aquesta població (la que es presenta en la taula 3.4.).

| CLASSIFICACIÓ | IMC (Kg/m ²) |
|-----------------|--------------------------|
| Normopès | 18,5 – 22,9 |
| Sobrepès | 23,00 – 24,9 |
| Obesitat | ≥ 25,00 |

Taula 3.4. Classificació dels IMC pels orientals.

També ens trobem amb problemes quan es tracta d'atletes, ja que l'IMC tendeix a considerar que persones sanes amb una massa muscular alta pareixen sobrepès. Per aquest col·lectiu no hi ha una solució tan clara; ja que l'IMC només depèn de l'alçada i el pes, i això provoca una tendència a sobreestimar l'adipositat en persones amb una major musculatura i/o ossada.

De totes maneres l'IMC és una mesura molt estesa i força emprada per conèixer si l'adipositat d'un individu és l'adequada. De fet no resulta gents estrany trobar balances electròniques a les farmàcies que també et mesuren i et diuen el teu IMC; com tampoc ho és veure anuncis per Internet de pàgines on et calculen el teu IMC, o determinen el teu pes ideal en funció de la teva alçada (lògicament aïllant el pes en el càlcul del IMC).

En el cas dels nens, existeix una altra aplicació de l'IMC molt coneguda. Consisteix en unes fitxes on hi ha representats uns determinats quantils de l'IMC segons l'edat, formant tot un seguit de corbes (figura 3.3). Mitjançant aquest gràfic es pretén elaborar una comparativa entre un individu i el valor típic per l'edat i les característiques del nen/a, permetent així als pediatres detectar possibles problemes o anomalies en el creixement i actuar en conseqüència.

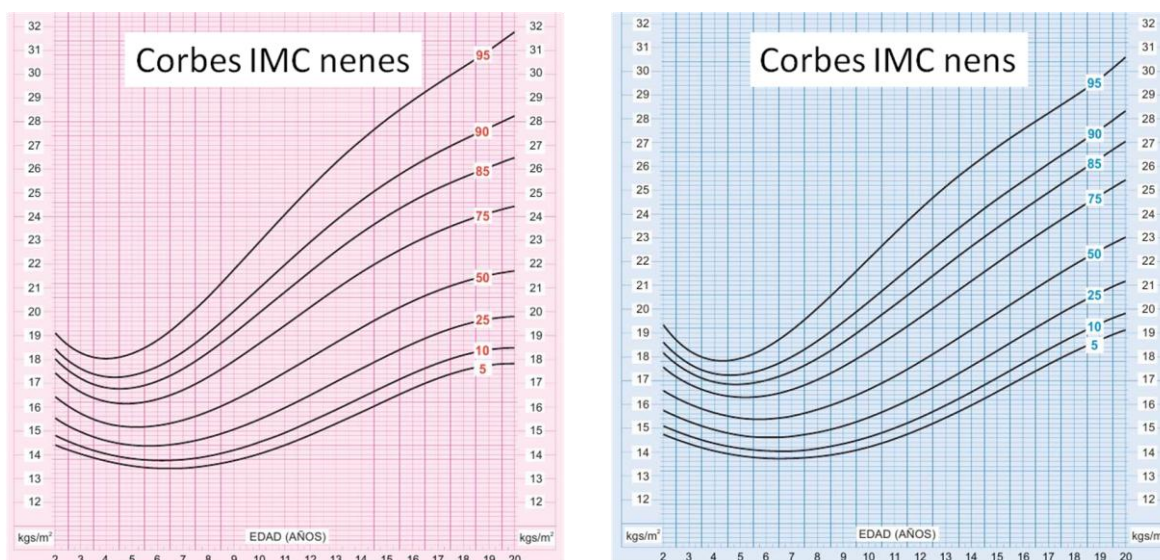


Figura 3.3. Corbes corresponents als quantils de l'IMC per a nens i nenes Espanyols. Font: <http://www.perseo.aesan.mspes.es> (Ministerio de Sanidad, Política social e Igualdad)

També existeixen uns gràfics iguals i emprats amb les mateixes finalitats, però on el que es representa no són valors de l'IMC sinó de l'alçada.

De fet, com ja s'ha comentat anteriorment, un dels objectius d'aquest projecte serà reproduir aquestes corbes per la població adulta; ja que no han estat trobades en la recerca bibliogràfica, corbes similars per a la població espanyola. El que volem és aconseguir una eina gràfica similar a aquesta (figura 3.3.) que ens permeti comparar als individus adults amb el comportament general, situant-los en el gràfic i veient prop de quin percentil cauen; tal i com els pediatres fan amb els seus pacients.

4. ANÀLISI DESCRIPTIVA

Per tal de poder tractar i manipular les dades, va caldre un important treball previ de comprensió i lectura d'aquestes. En la pàgina web oficial de l'INE es presenten en format ASCII a disposició de tothom, acompanyades de les corresponents metadades; així que obtenir-les fou senzill i directe.

La primera dificultat que se'ns va presentar va ser el tamany de les bases de dades. Aquestes tenien molts individus i nombroses variables (sobretot en el cas del 2006). Això feia impossible dur a terme la seva lectura amb programes com Minitab, que mostren les graelles per pantalla i posseeixen una mida màxima que no es pot sobrepassar. Per això, i donada la seva potència i la seva popularització, es va decidir treballar amb R.

La lectura es va encarar preparant unes taules amb excel on s'especificaven les variables i la longitud d'aquestes, ja que els fitxers disponibles no eren d'ample fix, per posteriorment passar-les al R juntament amb els arxius de les bases de dades. A priori semblava laboriós però senzill, donat que l'INE presentava uns documents amb les ENSE acompanyades de les metadades; on s'especificaven les variables i les seves longituds. Però un cop entrat tot tal i com els registres especificaven, les dues bases de dades (tant la del 1995 com la del 2006) presentaven errors de lectura. El principal problema era que, tot i que les dades es llegien correctament, els valors de les variables no concordaven amb el que indica l'arxiu de metadades.

Vist això, es dugué a terme un treball de revisió exhaustiva de les variables, de la base de dades en sí i dels registres; buscant algun possible error per part nostre o algun caràcter estrany en el document ASCII. Aleshores, comparant les ENSE amb els arxius corresponents als registres, es va descobrir que hi havia alguna variable recollida per l'enquesta i introduïda en la base de dades que no constava en els registres; així com una variable a la que se li havia donat una longitud diferent a l'especificada. Aquests eren els problemes que feien que la nostra lectura de dades fos errònia, i al detectar-los i modificar-los,

el codi ja funcionava correctament i els resultats obtinguts per a les variables coincidien amb els esperats (donat el coneixement previ i les especificacions dels documents de l'INE).

Després d'aquesta experiència, encara ens quedà més clar que una base de dades sempre és convenient acompanyar-la d'unes bones metadades, clares i sense errors; on es doni la informació necessària per comprendre la informació que posseïm i dur a terme una lectura correcta.

Un cop corregides, llegides i revisades les bases de dades, i previ a procedir amb les comparatives entre els dos anys i altres anàlisis; sempre és convenient realitzar una anàlisi descriptiva de les dades per fer-nos una idea del que hi passa i detectar possibles anomalies.

Cal destacar l'aparició de dades mancants (notificades com "NA's", tal i com ho fa el programa emprat). Per cada variable n'apareixen de diferents, ja que es tracta de valors que no coneixem perquè l'entrevistat no ha volgut donar aquesta informació; en la majoria dels casos.

Aparentment sembla que hi ha forces dades mancants, fet que seria preocupant si no fos perquè estem treballant amb bases de dades molt grans. Tot i això, no podem oblidar-nos de que hi són presents i ens poden estar ocultant algun fenomen; però el fet de que ens quedin molt casos ens permet seguir treballant amb un gruix d'informació important.

4.1. ANY 1995

La base de dades corresponent a l'any 1995 conté 64 variables avaluades en un total de 6396 individus. Aquestes variables corresponen bàsicament a l'àmbit mèdic i al socioeconòmic; i donat que provenen d'una enquesta, la majoria són categòriques o numèriques discretes (a excepció d'algunes, com el pes i l'alçada, que poden considerar-se contínues). Hi trobem variables relacionades amb els accidents i la salut personal, els hàbits de vida, el nivell d'estudis, els hàbits d'higiene i les característiques físiques i sensorials.

4.1.1. Depuració de la base de dades

Abans de començar amb l'anàlisi de les dades, procedirem a depurar-les per tal de procurar eliminar els possibles errors i/o dades anòmales que hi puguin aparèixer. Cal recordar que el que es pretén en aquest projecte es estimar com es comporta l'IMC en la població espanyola i comparar aquestes distribucions entre l'any 1995 i el 2006; així que necessitem eliminar aquells individus amb dades excessivament anòmales o amb valors erronis.

Per començar donarem un cop d'ull a l'edat per assegurar-nos que no hi ha errors i que tots els individus són adults (més de 16 anys).

- **Edat:**

| Min. | 1st Qu. | Median | Mean | 3rd Qu. | Max. | NA's | n |
|-------|---------|--------|-------|---------|-------|--------|------|
| 12.00 | 26.00 | 39.00 | 41.65 | 56.00 | 99.00 | 997.00 | 6396 |

Podem copsar que el mínim és 12, així que aquí hi ha com a mínim un individu que no hauria de ser a la nostra base de dades. Si ho consultem veiem que només és un nen de 12 anys. Com que no podem saber si es tracta d'un error en la transcripció de les dades o un nen realment, no té cap sentit mirar les característiques d'aquest individu. Al eliminar-lo tenim el següent:

| Min. | 1st Qu. | Median | Mean | 3rd Qu. | Max. | NA's | n |
|-------|---------|--------|-------|---------|-------|--------|------|
| 16.00 | 26.00 | 39.00 | 41.65 | 56.00 | 99.00 | 997.00 | 6395 |

Un cop ja hem depurat fixant-nos en l'edat, com que durant tot l'estudi ens centrarem en les variables del pes i l'alçada, creiem oportú observar la descriptiva numèrica d'aquestes dues variables. D'aquesta manera detectarem possibles individus anòmals que ens provocarien biaixos en estudis posteriors.

- **Alçada:**

| Min. | 1st Qu. | Median | Mean | 3rd Qu. | Max. | NA's | n |
|------|---------|--------|-------|---------|-------|-------|------|
| 48.0 | 160.0 | 166.0 | 166.4 | 172.0 | 204.0 | 841.0 | 6395 |

Pel que fa als màxims no sembla que hi hagi dades extremadament estranyes. En canvi, per la part dels mínims es detecta que hi ha algun problema donat que apareix un mínim de 48cm (valor molt estrany tractant-se de l'alçada d'un adult). Al mirar detalladament els individus amb valors baixos corresponents a l'alçada, observem que es tracta de més d'un cas (com es pot apreciar en la taula 4.1.).

| Sexe | Edat | Alçada (cm) | Pes (kg) | C. Autònoma |
|------|------|-------------|----------|-----------------|
| Dona | 56 | 58 | 84 | Madrid |
| Dona | 45 | 48 | 57 | Andalusia |
| Home | 40 | 75 | 170 | Catalunya |
| Home | 64 | 53 | 70 | Catalunya |
| Dona | 34 | 55 | 55 | Castella i Lleó |

Taula 4.1. Llistat d'individus amb valors anòmals per l'alçada a l'any 1995.

Donat que el rècord Guinness de la persona més baixa del món mesura 51cm, i que es tracta d'un fet molt estrany, creiem que els individus de la taula 4.1. poden ser errors de la base de dades. De totes maneres, encara que no fos

així, el que tenim clar és que aquests individus que mesuren menys de 80cm d'alçada no entren en el que és habitual. Vist tot això, i tenint en compte que en la base de dades inicial tenim 6.396 registres, decidim treure aquests 5 individus i seguir treballant amb els 6.391 restants. Ara l'alçada de l'any 1995 queda de la següent manera:

| Min. | 1st Qu. | Median | Mean | 3rd Qu. | Max. | NA's | n |
|-------|---------|--------|-------|---------|-------|-------|------|
| 132.0 | 160.0 | 166.0 | 166.5 | 172.0 | 204.0 | 841.0 | 6390 |

- **Pes:**

| Min. | 1st Qu. | Median | Mean | 3rd Qu. | Max. | NA's | n |
|-------|---------|--------|-------|---------|--------|--------|------|
| 35.00 | 60.00 | 69.00 | 69.54 | 78.00 | 695.00 | 991.00 | 6390 |

| Sexe | Edat | Alçada (cm) | Pes (kg) | C. Autònoma |
|------|------|-------------|----------|-------------|
| Home | 45 | 165 | 695 | Múrcia |

Taula 4.2. característiques de l'individu anòmal pel pes a l'any 1995.

Es pot observar que apareix un individu amb un pes de 695 kg. Al ser la base de dades del 1995 tenim molt poca informació, i això fa que no puguem saber si es tracta d'un error. De totes maneres, alguna de la poca informació que es té es presenta en la taula 4.2.. Sospitem que es pot tractar d'un error, i si no fos així es tracta d'un individu que no es troba dins del comportament habitual de la població. Vist això, decidim treure'l del nostre estudi perquè no enterboleixi els resultats de l'anàlisi posterior.

Per altra banda, copsem que a la base de dades també hi ha un individu que pesa 171kg i que el mínim està en 35kg. Observem les seves característiques

a la taula 4.3. i, tot i que són valors força extrems, decidim que poden ser “normals” i que no els exclourem de la nostra mostra.

| Sexe | Edat | Alçada (cm) | Pes (kg) | C. Autònoma |
|------|------|-------------|----------|---------------|
| Home | 30 | 180 | 171 | Múrcia |
| Dona | 75 | 158 | 35 | C. Valenciana |

Taula 4.3. Llistat d'individus amb valors extrems pel pes a l'any 1995.

Feta la depuració, la base de dades del 1995 queda amb 6.389 individus; mida que considerem prou gran i significativa per la nostra mostra.

4.1.2. Descriptiva univariant

Un cop depurada la base de dades, donem un cop d'ull als gràfics univariants per tenir una primera idea del que succeeix en la nostra base de dades.

- **Alçada:**

Com ja s'ha vist en l'apartat 3.2., l'alçada és un dels paràmetres amb els que es calcula l'IMC. Per aquesta raó, es converteix en una de les variables crucials pel nostre estudi; i procedim a analitzar-la individualment per tal d'obtenir informació útil per la comprensió de resultats posteriors.

| Min. | 1st Qu. | Median | Mean | St.Dev | 3rd Qu. | Max. | NA's | n |
|-------|---------|--------|-------|--------|---------|-------|-------|------|
| 132.0 | 160.0 | 166.0 | 166.5 | 9.96 | 172.0 | 204.0 | 841.0 | 6390 |

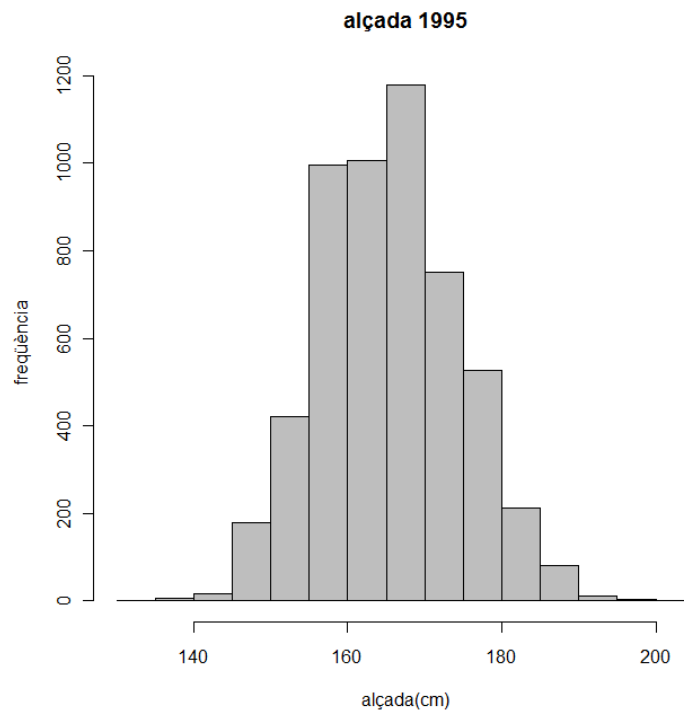


Figura 4.1. Histograma de l'alçada del 1995.

En la figura 4.1 veiem que l'histograma de l'alçada té una forma que s'assimila força a la d'una distribució normal. Això no és d'estranyar donada la naturalesa de la variable, i donat que veiem com un forat en el pic que esperàriem proper al valor central; procedim a representar aquestes dades en paper probabilístic normal per tenir un altre punt de vista.

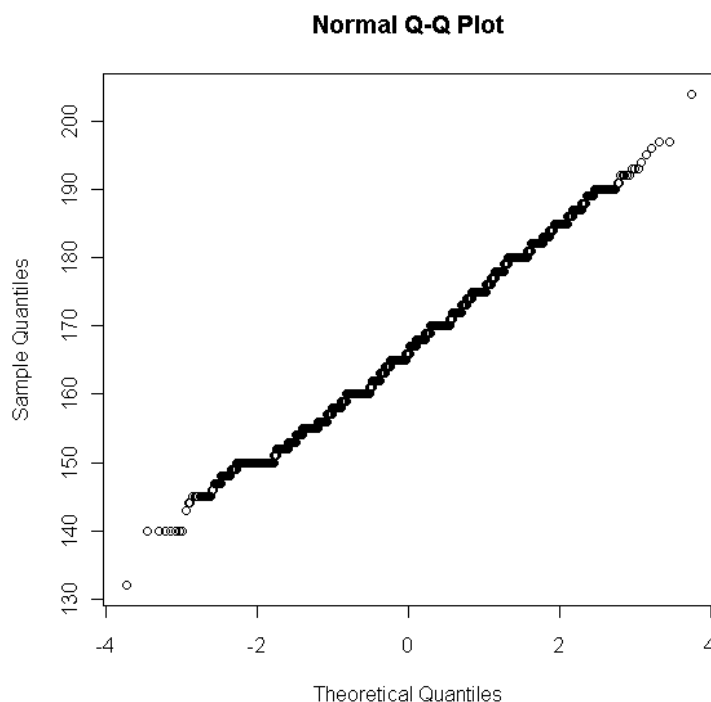


Figura 4.2. Dades de l'alçada del 1995 representades en paper probabilístic normal.

En la figura 4.2 cal destacar que les dades es troben totes bastant alineades amb la diagonal corresponent a la normalitat, però copsem que a les cues hi ha nombrosos valors anòmals. De totes maneres, no seria absurd creure que procedeixen d'una normal, i per això creiem convenient testar-ho mitjançant el test de "Kolmogorov-Smirnov".

Aquest test, que presentem formalment a continuació, consisteix en una prova no paramètrica de bondat d'ajust, així que el que ens determina és si és raonable creure que un conjunt procedeix d'una distribució determinada (en aquest cas la normal).

$$\begin{cases} H_0: \text{La mostra prové de la distribució proposada.} \\ H_1: \text{La mostra no prové de la distribució proposada.} \end{cases}$$

Estadístic de prova: $D_n = \sup_x |F_n(x) - F(x)|$, on $F_n(x)$ és la funció de distribució empírica calculada amb les dades que es testejen i $F(x)$ és la funció de distribució teòrica de la distribució normal (que és la proposada en aquest cas).

Per aplicar aquest test cal que les variables siguin contínues (com és el cas); i cal tenir en compte que, tot i tenir una potència alta i ser vàlid per a mostres de qualsevol mida, és un test que posseeix una alta tendència a rebutjar la hipòtesi nul·la (sobretot al treballar amb mostres grans).

Lilliefors (Kolmogorov-Smirnov) normality test

data: dades95\$alçada

D = 0.0674, p-value < 2.2e-16

El p.valor obtingut és molt més petit que 0,05, fet que ens porta a refusar la hipòtesi nul·la al treballar amb $\alpha=0,05$. Així que estrictament parlant no es pot considerar que aquesta mostra procedeixi d'una distribució normal; i això no és d'estranyar per molt que ens ho sembli, ja que donat que tenim moltes dades és realment difícil que al passar el test de bondat d'ajust no rebutgem H_0 .

- **Pes:**

Aquesta és, juntament amb l'alçada, l'altra variable que conforma l'IMC. Per aquesta raó adopta un paper molt important en tots les anàlisis sobre l'adipositat que elaborarem posteriorment; fet que ens fa creure necessari donar-hi un cop d'ull per extreure'n informació.

| Min. | 1st Qu. | Median | Mean | St.Dev | 3rd Qu. | Max. | NA's | n |
|-------|---------|--------|-------|--------|---------|--------|--------|------|
| 35.00 | 60.00 | 69.00 | 69.44 | 12.9 | 78.00 | 171.00 | 997.00 | 6389 |

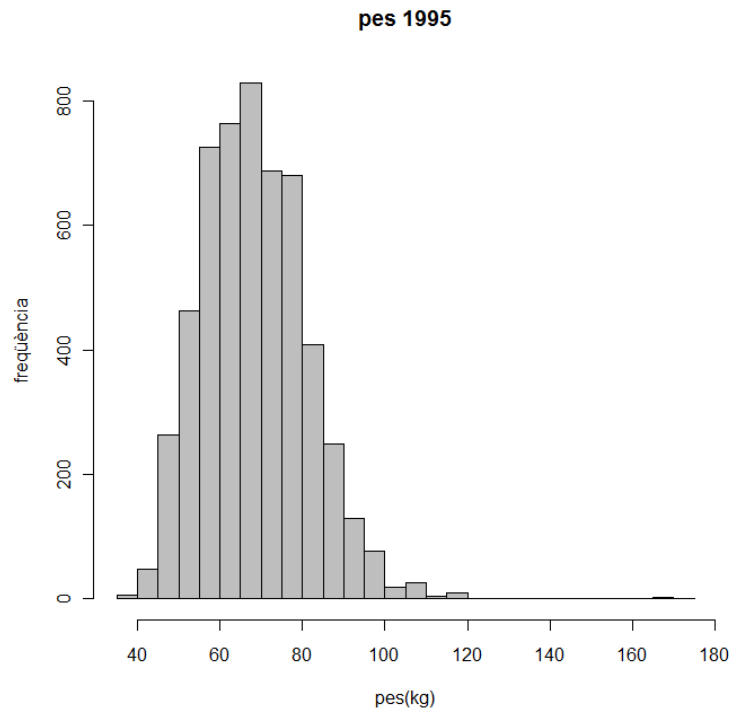


Figura 4.3. Histograma del pes del 1995.

L'histograma corresponent al pes del 1995 (Fig. 4.3) presenta una censura inferior, fet que no és d'estranyar donat que la nostra mostra correspon a la població adulta. El pes mínim observat és de 35 quilos, i ja es pot copsar que resultaria difícil un pes inferior a aquest donada la complexió del cos humà.

L'histograma presenta una forma semblant a una campana de Gauss però amb una clara cua a la dreta més pesada que en el cas de la normal. Aquesta, sumada al tall en la part baixa de la distribució, fa que sembli força clar que aquestes dades no procedeixen d'una distribució normal. Tot i així es pot passar un test de bondat d'ajust per veure que, efectivament, aquestes dades no s'ajusten a una distribució normal.

Lilliefors (Kolmogorov-Smirnov) normality test

data: dades95\$pes

D = 0.0573, p-value < 2.2e-16

- **IMC:**

A partir de les dues variables presentades anteriorment, s'ha calculat l'IMC (tal i com s'especifica en l'apartat 3.2.). Mitjançant aquest índex pretenem capturar la informació sobre l'adipositat dels individus de les nostres mostres, permetent així elaborar les comparatives i les anàlisis que ens hem proposat.

Abans, però, de procedir amb aquests estudis creiem necessari abordar un anàlisi univariant del IMC per copsar-ne el seu comportament en la base de dades corresponent al 1995.

| Min. | 1st Qu. | Median | Mean | St.Dev | 3rd Qu. | Max. | NA's | n |
|-------|---------|--------|-------|--------|---------|-------|--------|------|
| 14.02 | 22.13 | 24.62 | 25.00 | 4.07 | 27.34 | 59.88 | 991.00 | 6389 |

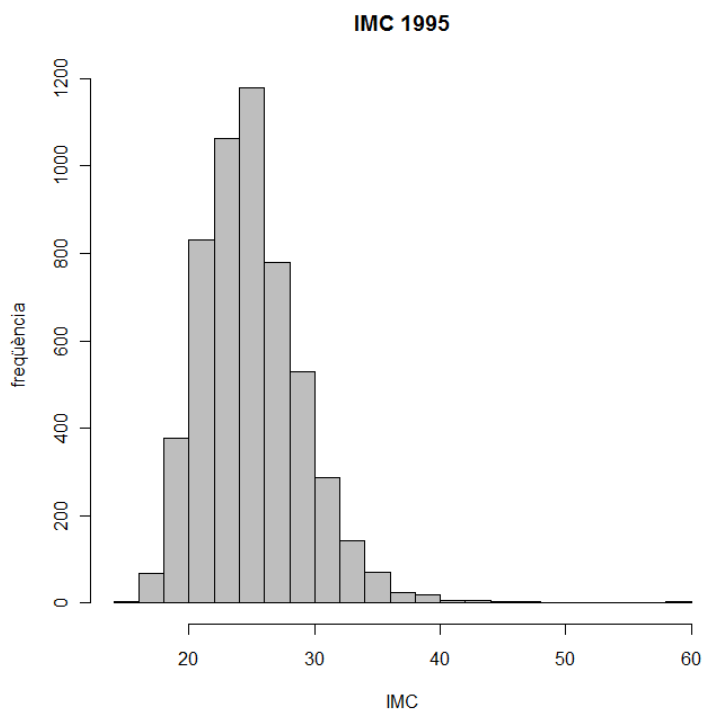


Figura 4.4. Histograma del IMC de l'any 1995.

En l'histograma de l'IMC (figura 4.4.) observem quelcom similar al que acabem de veure en l'histograma del pes (figura 4.3.). Està censurat per sota, ja que a partir d'un determinat valor ja no es poden trobar individus amb un IMC més petit. També es pot copsar que presenta una cua dreta força gran en comparació amb la distribució normal. No és d'estranyar que presenti aquesta forma lleugerament asimètrica assimilant-se al comportament del pes, ja que aquest intervé en el càlcul de l'IMC (veure apartat 3.2.).

Tal i com passava amb el pes, donada la representació gràfica ja no es creu que aquestes dades provinguin d'una distribució normal i així ho manifesta el test de Kolmogorov-Smirnov que, amb un p.valor petitíssim, en duu a refusar la hipòtesis nul·la.

Lilliefors (Kolmogorov-Smirnov) normality test

data: dades95\$IMC

D = 0.0578, p-value < 2.2e-16

- **IMC categoritzat:**

Mitjançant l'IMC anterior, s'ha creat la variable categòrica corresponent amb els grups indicats per a la categorització estàndard d'aquest índex (taula 3.3.).

Procedim a analitzar-la per fer-nos una primera idea del que passa en la base de dades del 1995, pel que fa als grups corresponents.

| Infrapes | Normopes | Sobrepes | Obesitat |
|----------|----------|----------|----------|
| 130 | 2831 | 1872 | 566 |
| 2,69% | 58,58% | 38,73% | 11,71% |

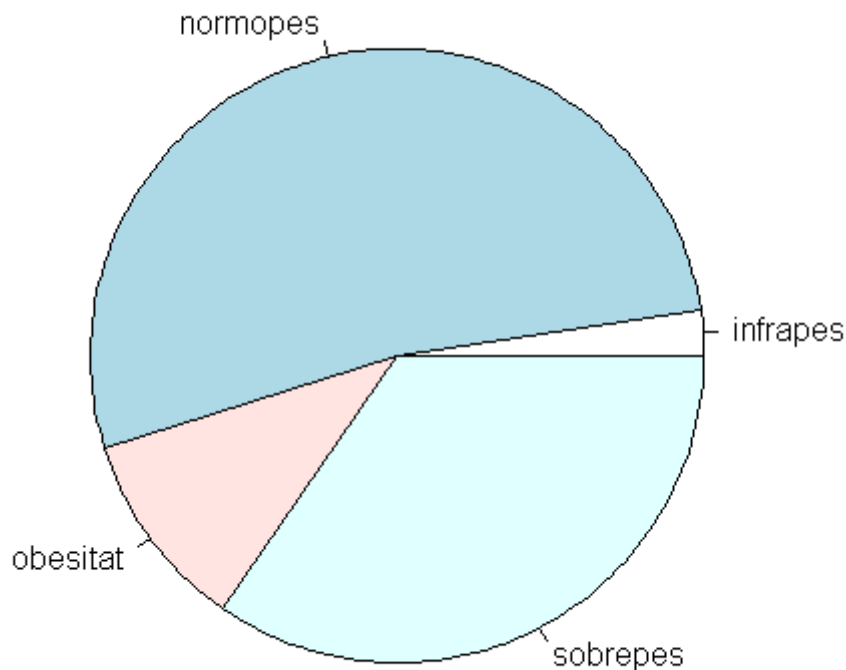


Figura 4.5. Diagrama de sectors del IMC categoritzat de l'any 1995.

Gairebé la meitat de la nostra mostra presenta normopès (figura 4.5.), molt bona senyal donat que el millor sempre és trobar-se en aquest grup (no estar ni massa prim ni massa gras). Respecte la resta dels individus, s'observa que n'hi ha molt pocs al grup de l'infrapès. La majoria dels que es troben fora del grup ideal, es situen doncs en els grups del sobrepès i la obesitat. Aquest fenomen ens indica que sembla que la gent tendeix més a presentar una adipositat major a la convenient, que no pas menor.

Un cop vist això ens sembla força interessant donar un cop d'ull al comportament de la variable IMC numèrica dins de cadascun dels grups, avaluant-ne la mitjana i la desviació tipus. D'aquesta manera podrem copsar si la variabilitat interna de cadascun d'aquest grups s'assimila entre ells o presenta diferències.

També ens permetrà copsar la tendència central de cada grup, sent conscients que es veu afectada per la categorització emprada. Tot i això, ens pot ajudar a

veure si dins dels grups es tendeix més cap a valors relativament alts o baixos per aquella categoria.

| | Mean | St. Dev. |
|-----------------|--------|----------|
| infrapes | 17.776 | 0.738 |
| normopes | 22.364 | 1.717 |
| sobrepes | 27.064 | 1.369 |
| obesitat | 32.984 | 3.490 |

Podem veure com el grup de les persones amb una adipositat major presenta una dispersió notablement més gran. Així doncs, sembla que els més grassos varien més els seus valors de IMC. El que passa és aquesta categoria no té límit superior, i això ens podria ajudar a entendre una mica més el perquè d'aquesta alta dispersió. Tot i això, és preocupant descobrir aquesta elevada variabilitat, ja que ens fa veure un cop més que dins el grup dels més grassos hi ha individus amb valors d'adipositat altíssims.

El grup de l'infrapès presenta una variabilitat destacablement menor a la resta, quelcom força raonable si es té en compte que aquest és el grup que conté la menor part dels nostre individus (només el 2,4% de la mostra). De totes maneres també denota que no es tendeix a valors tant extremadament baixos d'adipositat com alts, vista la mitjana que presenta el grup de l'infrapès.

Pel que fa als grups del normopès i del sobrepès, veiem que presenten variabilitats força semblants. També cal destacar que el grup del normopès (on es troben els individus amb valors d'adipositat ideals), té una mitjana per sobre del centre dels valors que conformen aquest grup (21,75). Aquest

esdeveniment ens denota que dins dels individus amb adipositats adequades, hi ha una major tendència a apropar-se al sobrepès que no pas a l'infrapès.

Per tal de saber si totes aquestes diferències en la dispersió són estadísticament significatives, s'ha elaborat el test de Levene:

$$\begin{cases} H_0: \text{Hi ha homogeneïtat de variàncies entre els diferents grups.} \\ H_1: \text{No hi ha homogeneïtat de variàncies entre els grups.} \end{cases}$$

Per aquest test s'utilitza un estadístic F de Fisher-Snedecor corregit, donat que les variables no compleixen la premissa de normalitat.

Al elaborar-lo s'obté un p.valor inferior a $2,2 \cdot 10^{-16}$, fet que ens indica que les variàncies destacades anteriorment presenten diferències significatives; ja que no es pot considerar que hi hagi homogeneïtat entre els grups.

- **Edat:**

Donat que en posteriors anàlisis es crearà una categorització per l'edat i s'estratificaran els resultats per aquesta; val la pena fer una ullada a la distribució de les edats a la mostra del 1995.

| Min. | 1st Qu. | Median | Mean | St.Dev | 3rd Qu. | Max. | NA's | n |
|-------|---------|--------|-------|--------|---------|-------|--------|------|
| 16.00 | 26.00 | 39.00 | 41.65 | 18 | 56.00 | 99.00 | 997.00 | 6395 |

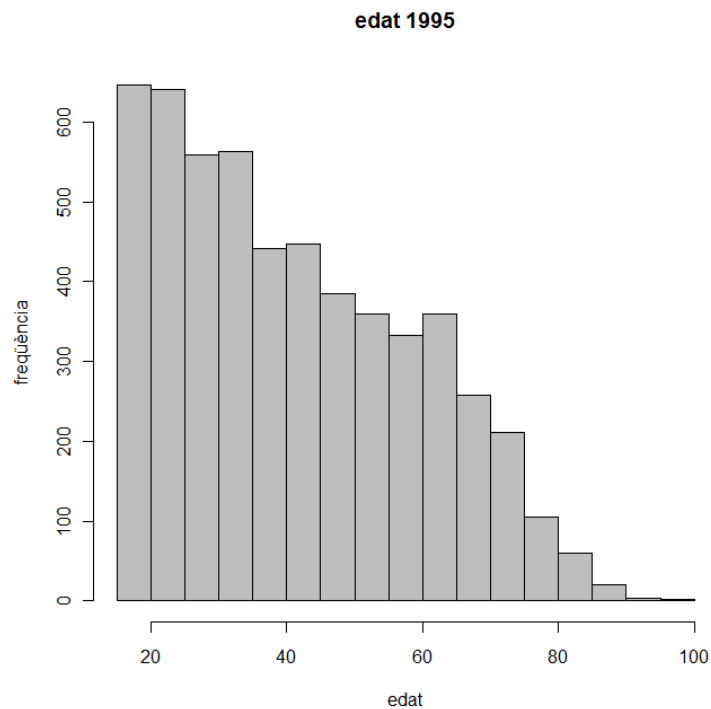


Figura 4.6. Histograma de l'edat de la mostra del 1995.

L'histograma de l'edat (figura 4.6.) evidentment està censurat, ja que la nostra mostra és de persones de 16 anys o més. A més observem un comportament que ens recorda una mica a la bimodalitat, fet que podria estranyar-nos i fer-nos pensar en la necessitat d'estratificar per alguna variable (com podria ser el sexe). El que farem, però, és observar la piràmide poblacional esdevinguda del cens d'aquest any (del 1995) per veure si la forma del nostre histograma s'adequa al que s'espera.



Figura 4.7. Piràmide poblacional espanyola corresponent a l'any 1995. Font: INE, cens de l'1 de gener del 1995.

Si mirem aquesta piràmide poblacional (figura 4.7.) i ens fem la idea de com seria el sobreposar els dos costats d'aquesta, corresponents als diferents sexes, veiem que la forma concorda força amb el que havíem observat en l'histograma de la nostra mostra (figura 4.6.).

- **Sexe:**

Una última variable a treballar de manera individual, és el sexe. Aquesta ens serà molt útil per a estratificar en les nostres anàlisis, així que val la pena obtenir-ne informació.

| Dona | Home |
|------|------|
| 2604 | 2794 |

Pel que fa a la variable del sexe, podem observar que gairebé la meitat de la nostra mostra són homes (51,8%) i la meitat dones (48,2%). Aquestes proporcions s'aproximen al 50%; però veiem que en la mostra corresponent al 1995 hi ha un percentatge major d'homes que de dones, i això és justament el contrari del que succeeix en la població.

Això no vol dir que la mostra sigui errònia, de fet ja es sabia que ens podíem trobar en aquesta situació donada la component d'aleatorietat de la mostra, però cal tenir-ho en compte.

4.1.3. Descriptiva bivariant

En aquest apartat es pretén observar algunes relacions entre les variables presentades anteriorment.

- **Alçada VS pes:**

En primer lloc es presenta la relació entre l'alçada i el pes dels individus de la mostra. És força intuïtiu creure que poden presentar alguna relació, així que procedim donar-hi un cop d'ull.

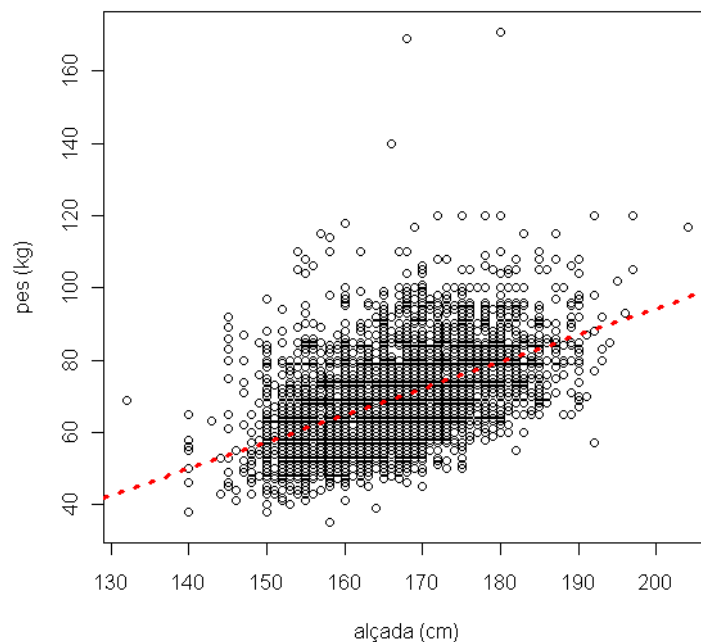


Figura 4.8. Diagrama de punts on es representen els valors dels individus per l'alçada i el pes en el 1995.

Veient el diagrama de punts (figura 4.8.) es pot veure que existeix una relació significativa entre aquestes dues variables; on a major alçada, més pes. Aquesta relació és força natural, ja que acostuma a donar-se que les persones més altes pesen més que les baixetes donada la seva complexió. Amb això no es vol dir que una persona més alta sempre pesi més que una altra més baixa, però és natural que persones amb una alçada més gran presentin una complexió major i per tant un pes també més gran.

De totes maneres la recta estimada explica poc la relació entre ambdues variables donada la gran variabilitat de les dades; fet que ens indica que, tot hi haver-hi relació, cada individu actua d'una manera força diferent en quant a aquesta.

- **Edat VS IMC:**

Donada la importància de l'IMC en el nostre estudi, ens sembla força interessant veure si existeix relació entre la edat i aquest índex.

Aquesta relació ja no és tant natural com l'efectuada anteriorment, i menys tractant-se d'una població únicament adulta. Aquesta relació seria una primera aproximació a quelcom com respondre a la pregunta: a mida que ens anem fent grans som més grassos?

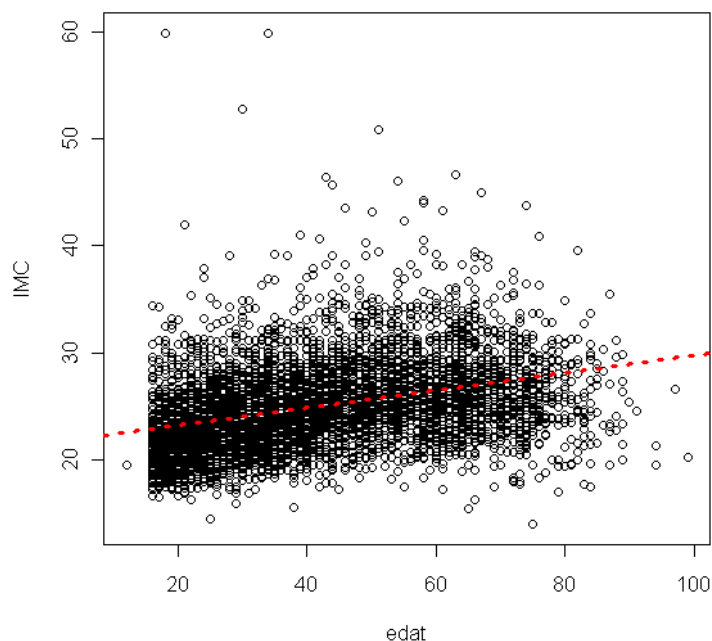


Figura 4.9. Diagrama de punts on es representen els valors dels individus per l'edat i l'IMC en el 1995.

Es pot observar a la figura 4.9. una certa relació estadísticament significativa. Això ens duria a creure que, a grans trets, a mida que passen els anys va augmentant la nostra adipositat. De totes maneres aquesta relació és molt lleugera, així que no podem dir que l'edat expliqui massa l'IMC.

- **IMC estratificant per sexe:**

Donada la importància de l'adipositat en el nostre estudi, i sent que per la població estàndard l'IMC és la millor manera de mesurar-la, volem comprendre bé el comportament d'aquest índex en la nostra base de dades.

Sabem que, en termes generals, homes i dones presenten complexions diferents; així que creiem interessant veure les diferències entre sexes per l'IMC.

| | n | Mean | St. Dev. | Min. | 1st Qu. | Median | 3rd Qu. | Max. |
|-------------|------|--------|----------|-------|---------|--------|---------|-------|
| Dona | 2599 | 24.445 | 4.453 | 14.02 | 21.23 | 23.62 | 26.93 | 59.88 |
| Home | 2792 | 25.517 | 3.600 | 15.46 | 23.15 | 25.21 | 27.64 | 59.88 |

Numèricament destaquen les diferències entre les mitjanes i les desviacions tipus. Per veure si aquestes diferències són estadísticament significatives, elaborarem un test amb l'estadístic F de Fisher-Snedecor per les variàncies i un de la t d'Student. Aquests test suposen normalitat, però donat el gran nombre de dades creiem que és una aproximació vàlida.

El test d'igualtat de variàncies és el següent:

$$\begin{cases} H_0: \text{Les variàncies de cada grup són estadísticament diferents.} \\ H_1: \text{Les variàncies no són estadísticament diferents.} \end{cases}$$

Al fer-lo es rebutja clarament la hipòtesis nul·la, donat que s'obté un p.valor inferior a $2,2 \cdot 10^{-16}$.

Aleshores procedim a elaborar el test mitjançant una t d'Student, però cal tenir en compte que no podem suposar que les variàncies siguin iguals (donat que acabem de veure que no és així):

$$\begin{cases} H_0: \text{La mitjana és igual per ambdós grups.} \\ H_1: \text{La mitjana no és igual per ambdós grups.} \end{cases}$$

En aquest cas també obtenim un p.valor inferior a $2,2 \cdot 10^{-16}$, la qual cosa ens duu a refusar la hipòtesis nul·la; provant així una diferència estadísticament significativa entre la mitjana de l'IMC per a dones i per a homes.

Així que el que observem és que les dones presenten una mitjana per sota de la dels homes, però aquests darrers presenten una dispersió menor que les dones en quant a l'adipositat. Si mirem més a fons la descriptiva numèrica, concloem que aquesta diferència de dispersió prové del fet que les dones arriben a valors més baixos que els homes, i en canvi ambdós sexes comparteixen màxim. Això provoca que el rang de les dones sigui major al dels homes, i ens ajuda a comprendre el perquè de les diferències en les desviacions tipus.

La figura 4.10. corrobora de manera gràfica el que s'ha comentat anteriorment. A més també copsem que els homes presenten una distribució més simètrica que les dones. Això es pot apreciar gràficament, però sobretot es destaca en la similitud entre la mediana i la mitjana en el cas dels homes; fet que no es dona en el cas de les dones.

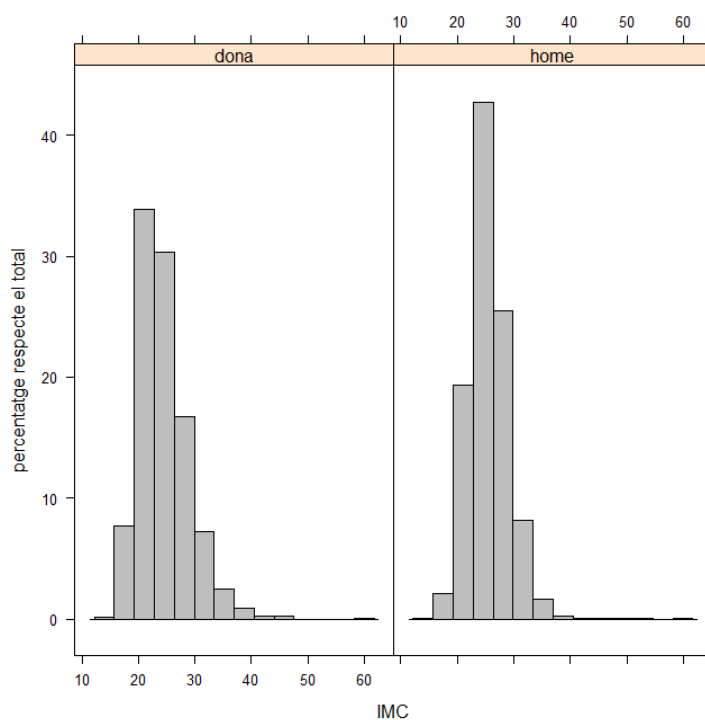


Figura 4.10. Histogrames de l'IMC del 1995 estratificant per sexe.

4.2. ANY 2006

Un cop analitzada la base de dades del 1995, procedirem anàlogament per les dades del 2006.

4.2.1. Depuració de la base de dades

La base de dades de l'any 2006 consta de 619 variables avaluades en un total de 29.478 individus. En aquestes cas tenim un major nombre de variables que en el del 1995, i una major varietat d'aspectes tractats; així que es tracta d'una base de dades molt més completa que la del 1995 (com es pot veure si es comparen les ENSE de l'annex 9.3. i 9.4.).

Les variables que hi trobem responen a preguntes socioeconòmiques, del treball reproductiu, salut, hàbits de vida, agressions, estrès laboral, pràctiques preventives, discriminació característiques físiques i sensorials, suport afectiu i personal i funció familiar; entre d'altres. Tal i com passava amb la base de dades del 1995, la majoria són categòriques o numèriques discretes, ja que provenen de respostes d'una enquesta (a excepció d'algunes, com el pes i l'alçada, que poden considerar-se contínues). De totes maneres ja es pot veure que posseïm molta més informació sobre aquesta mostra; que ha respost una enquesta en la que es tracten temes molt més diversos (annex 9.4.).

Com ja hem fet amb l'any 1995, començarem assegurant-nos que la nostra base de dades està formada només per individus adults.

- **Edat:**

| Min. | 1st Qu. | Median | Mean | 3rd Qu. | Max. | NA's | n |
|-------|---------|--------|-------|---------|--------|------|-------|
| 16.00 | 36.00 | 49.00 | 50.61 | 66.00 | 104.00 | 0 | 29478 |

Observant la descriptiva decidim no treure cap valor, ja que tots són majors de 16 anys. De totes maneres pot ser que per algunes parts del nostre estudi decidim tallar la mostra per dalt per evitar treballar amb edats molt elevades que presenten pocs individus.

A continuació, donarem un cop d'ull a l'alçada i el pes pel mateix motiu que en el cas del 1995.

- **Alçada:**

| Min. | 1st Qu. | Median | Mean | 3rd Qu. | Max. | NA's | n |
|------|---------|--------|-------|---------|-------|--------|-------|
| 57.0 | 160.0 | 165.0 | 165.5 | 172.0 | 203.0 | 2716.0 | 29478 |

Seguint el mateix criteri que amb les dades de l'any 1995, eliminarem de l'estudi els individus l'alçada dels quals es trobi per sota dels 80cm (taula 4.4.). En aquest cas només n'hi ha un (el qual mesura 57cm) i per tant, ja que inicialment comptàvem amb 29.478 individus, continuem l'estudi amb 29.477.

| Sexe | Edat | Alçada (cm) | Pes (kg) | C. Autònoma |
|------|------|-------------|----------|-------------|
| Dona | 37 | 57 | * | Catalunya |

Taula 4.4. Característiques de l'individu de la mostra del 2006 eliminat per la seva baixa estatura.

Ara, l'alçada de l'any 2006 queda de la següent manera:

| Min. | 1st Qu. | Median | Mean | 3rd Qu. | Max. | NA's | n |
|-------|---------|--------|-------|---------|-------|--------|-------|
| 105.0 | 160.0 | 165.0 | 165.5 | 172.0 | 203.0 | 2716.0 | 29477 |

- **Pes:**

| Min. | 1st Qu. | Median | Mean | 3rd Qu. | Max. | NA's | n |
|------|---------|--------|------|---------|-------|--------|-------|
| 25.0 | 60.0 | 70.0 | 70.9 | 80.0 | 180.0 | 3273.0 | 29477 |

S'observa un valor mínim molt baix així que decidim observar-lo més a fons. A més, aprofitarem que la base de dades de l'any 2006 conté més informació per mirar aquest individu amb detall abans de prendre una decisió (taula 4.5.).

| Sexe | Edat | Alçada (cm) | Pes (kg) | C. Autònoma | Malalties |
|------|------|-------------|----------|----------------|--|
| Dona | 19 | 140 | 25 | P. de Astúries | Atacs d'epilèpsia, hernies discals i tetraplegia |

Taula 4.5. característiques de l'individu amb un valor anòmal pel pes en el 2006.

Per les malalties que presenta, sembla que aquesta dona està postrada en un llit i que no pot mantenir una vida normal. A més és molt jove i això fa que, el fet que presenti tants problemes de salut, és molt rar. Per tots aquests motius creiem que aquest individu és anòmal i per això el treiem de l'estudi. Aleshores la variable pes de l'any 2006 queda de la següent manera:

| Min. | 1st Qu. | Median | Mean | 3rd Qu. | Max. | NA's | n |
|------|---------|--------|------|---------|-------|--------|------|
| 30.0 | 60.0 | 70.0 | 70.9 | 80.0 | 180.0 | 3273.0 | 2976 |

Es pot observar que ara apareix un individu que pesa 30kg. Aquest no és un pes que sembli extremadament estrany, però de totes maneres donarem un cop d'ull a aquest individu per assegurar-nos de que no es tracta d'una anomalia (taula 4.6.).

| Sexe | Edat | Alçada (cm) | Pes (kg) | C. Autònoma | Malalties |
|------|------|-------------|----------|-------------|--|
| Dona | 84 | 140 | 30 | Galícia | Hernies discals, colon irritable i perforació del timpà. |

Taula 4.6. Característiques de l'individu del 2006 que ens crida l'atenció donat el seu pes.

Podríem dir que es tracta d'un individu que està dins del comportament habitual de la població, donada la seva avançada edat. Tenim alguns dubtes però decidim conservar-la en la base de dades del nostre estudi.

Feta la depuració, la base de dades del 2006 queda amb 29.476 individus; mida més que suficientment gran i significativa per la nostra mostra.

4.2.2. Descriptiva univariant

- **Alçada:**

Tal i com ja hem comentat, ens interessa especialment donar un cop d'ull a l'alçada per ser que forma par del càlcul de l'IMC.

| Min. | 1st Qu. | Median | Mean | St.Dev | 3rd Qu. | Max. | NA's | n |
|-------|---------|--------|-------|--------|---------|-------|--------|-------|
| 105.0 | 160.0 | 165.0 | 165.5 | 9.25 | 172.0 | 203.0 | 2716.0 | 29477 |

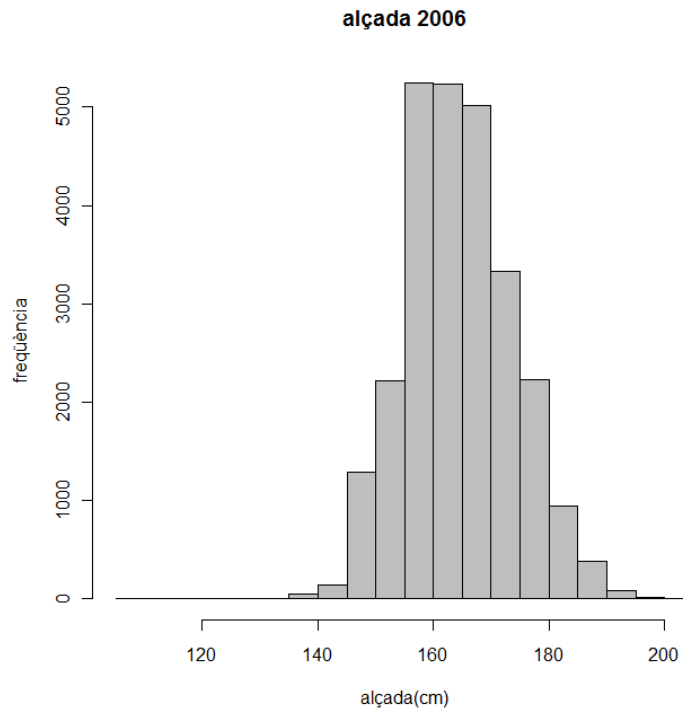


Figura 4.11. Histograma de l'alçada corresponent a la mostra del 2006.

En l'histograma de l'alçada (figura 4.11.) es torna a observar una similitud destacable amb la forma d'una distribució normal, tal i com ja passava amb l'alçada corresponent a la mostra del 1995 (figura 4.1.). La forma, en aquest cas, és lleugerament diferent; donat que en comptes d'un forat en el pic central, sembla que hi ha massa individus propers a aquella zona. També es pot copsar una lleugera inclinació cap a la dreta, ja que en aquella zona sembla que hi ha més individus que tendeixen cap al valor mig i menys cap a una cua que sembla acabar-se d'una manera una mica dràstica.

Donada l'escala de l'histograma, sembla que hi hagi un forat en la part corresponent als valors baixos. Això es dona perquè, tal i com hem observat en la depuració, hi ha algun individu força baixet que hem mantingut en l'estudi i que està en aquella zona.

Tot plegat, unit a que en la base de dades del 1995 semblava més una campana de Gauss i hem vist que no podíem acceptar que les dades corresponguessin a una distribució normal, sembla fer-nos dubtar força d'una

normalitat que no podem acceptar donat el p.valor del test presentat a continuació.

Lilliefors (Kolmogorov-Smirnov) normality test

data: dades06\$alçada

D = 0.066, p-value < 2.2e-16

- **Pes:**

També en aquest cas ens convé analitzar individualment la variable del pes, que com ja sabem és una de les variables que emprem pel càlcul de la variable que concentra el màxim interès d'aquest projecte; l'IMC.

Amb l'objectiu de veure com es comporta el pes en la mostra del 2006 procedim al seu anàlisi univariant.

| Min. | 1st Qu. | Median | Mean | St.Dev | 3rd Qu. | Max. | NA's | n |
|------|---------|--------|------|--------|---------|-------|--------|------|
| 30.0 | 60.0 | 70.0 | 70.9 | 13.96 | 80.0 | 180.0 | 3273.0 | 2976 |

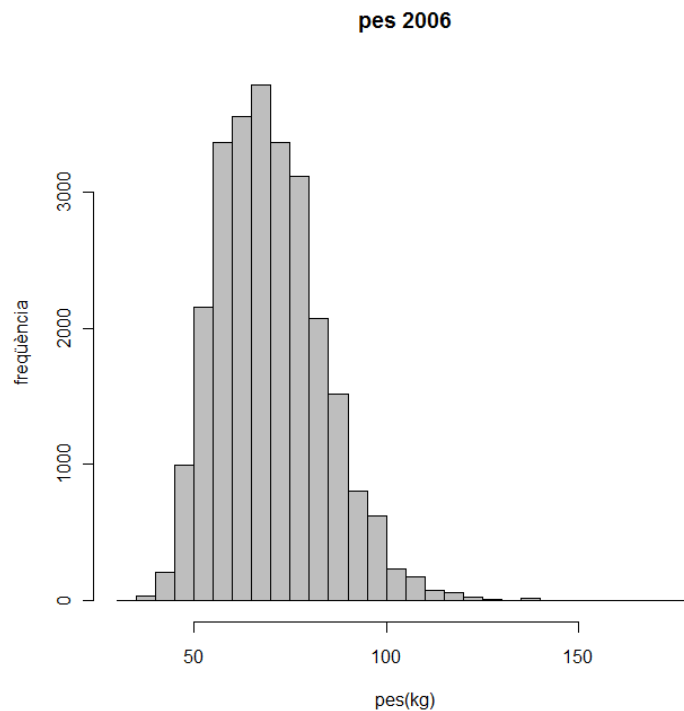


Figura 4.12. Representació gràfica del pes de l'any 2006.

La variable corresponent al pes de cada individu presenta una censura inferior, tal i com passava en la base de dades del 1995. De la mateixa manera que aleshores, atribuïm aquest fet a que estem tractant poblacions adultes i tot adult ha de pesar un mínim donada la composició del cos humà. Tot i no saber aquest límit, sabem que aquesta és una de les naturaleses habituals de les censures i concorda amb la realitat del cas.

A més d'aquesta censura, tornem a copsar una cua notablement més pesada que en el cas de la normal en els valors alts de la distribució. Això provoca aquest descentrament visible en la figura 4.12.; i tot plegat ens fa creure que, tal i com passava en la mostra del 1995, aquestes dades no provenen d'una distribució normal. El test de Kolmogorov-Smirnov així ens ho indica.

Lilliefors (Kolmogorov-Smirnov) normality test

data: dades06\$pes

D = 0.0641, p-value < 2.2e-16

- **IMC:**

Com ja s'ha comentat, aquesta és la variable que centra un major interès en la totalitat de la nostra anàlisi. Per això és clar que ens interessa tota la informació que en puguem extreure, començant per l'anàlisi univariant presentat a continuació.

| Min. | 1st Qu. | Median | Mean | St.Dev | 3rd Qu. | Max. | NA's | n |
|-------|---------|--------|-------|--------|---------|-------|---------|-------|
| 13.96 | 22.76 | 25.39 | 25.84 | 4.44 | 28.37 | 58.96 | 3273.00 | 29476 |

A les vistes de representació gràfica (figura 4.13.), podem dir que es tracta d'una variable que presenta una cua superior pesada; tal i com ho feia el pes i l'IMC del 1995.

La descriptiva numèrica ens sembla indicar que es tracta d'una variable força simètrica, donada la similitud entre la mitjana i la mediana; però veient l'histograma (figura 4.13.) i aquesta cua superior, veiem que no acaba de ser simètrica.

Tot i que no se'n pot extreure conclusions, si comparem aquesta descriptiva amb la del 1995 veiem que el màxim del 2006 és menor. També copsem que aquesta presenta un mínim més petit que el que tenia la base de dades del 1995, però per contra els valors corresponents a la mitjana i la mediana són majors en el cas del 2006.

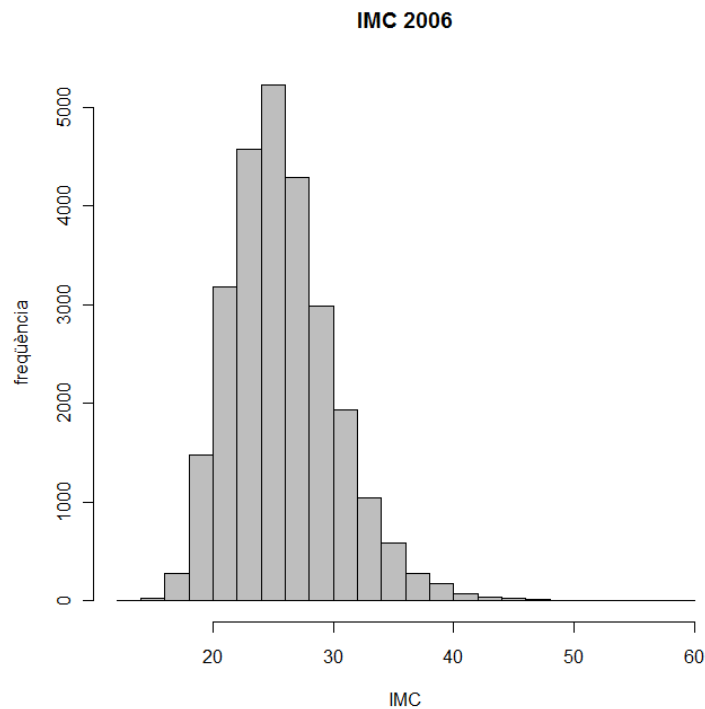


Figura 4.13. Representació gràfica del IMC corresponent a la mostra del 2006.

La possible censura inferior i sobretot la cua superior, ens torna a fer creure un cop més que aquesta variable no s'ajusta a una campana de Gauss. Veient l'histograma de la variable (figura 4.13.), ja es fa difícil creure una relació amb la distribució normal que el test de bondat d'ajust rebutja clarament; tal i com es pot veure en el test presentat tot seguit.

Lilliefors (Kolmogorov-Smirnov) normality test

data: dades06\$IMC

D = 0.0524, p-value < 2.2e-16

- **IMC categoritzat:**

Mitjançant la variable de l'IMC, calculada a partir del pes i l'alçada; s'ha creat aquesta categorització segons les indicacions de la taula de correspondències estàndard (taula 3.3.). Aquesta ens sembla una variable molt interessant donada la seva popularitat en la societat, i la gran informació sobre l'adipositat que ens pot proporcionar.

| Infrapès | Normopès | Sobrepès | Obesitat |
|----------|----------|----------|----------|
| 485 | 11821 | 9719 | 4178 |
| 1,85% | 45,11% | 37,09% | 15,95% |

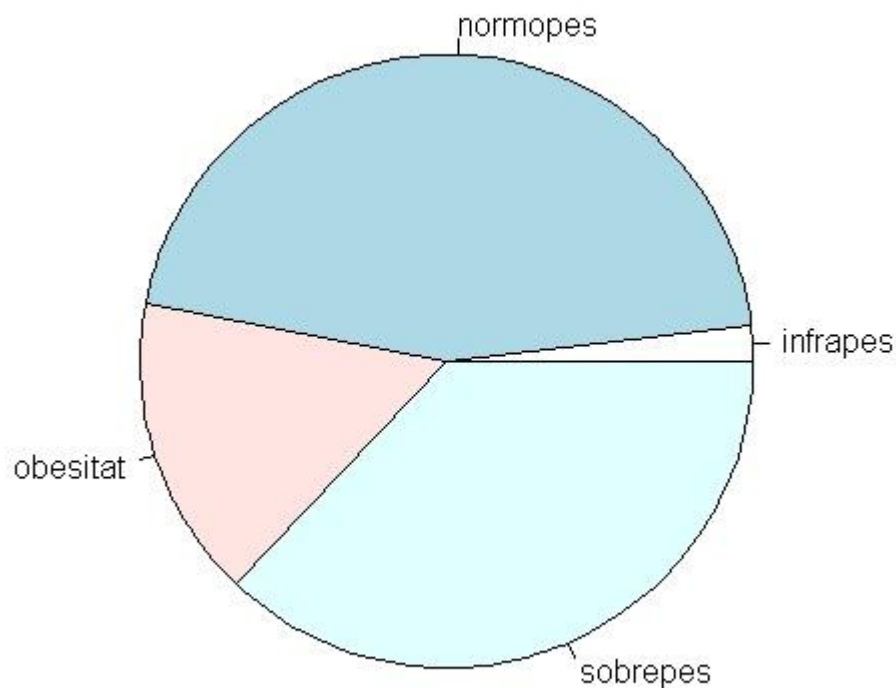


Figura 4.14. Diagrama de sectors on es representa l'IMC del 2006 per categories.

Observant la taula on es presenten els casos en cada grup, i el diagrama de sectors (figura 4.14.); és clar que els grups que contenen un major percentatge de la nostra mostra són els del normopès i el sobrepès.

Pel que fa al grup de la obesitat, es pot copsar que agafa un percentatge major de la població en aquesta mostra que en la del 1995; ja que ara es troba per damunt del 15% i en l'altre cas contenia poc més del 10% de la mostra. També s'observa que, tal i com passava amb la mostra del 1995, el grup amb menys individus és el de l'infrapès; que en aquest cas encara en conté un percentatge menor que en l'altra mostra.

Tot plegat ens indica que s'ha produït un increment de la proporció de persones en els grups corresponents a adipositats elevades; enfront d'un decrement de la proporció d'individus en els grups amb adipositats baixes.

Tal i com hem fet amb les dades del 1995, observem la mitjana i la desviació de l'IMC dins de cada categoria.

| | Mean | St. Dev. |
|-----------------|--------|----------|
| infrapes | 17.580 | 0.820 |
| normopes | 22.450 | 1.718 |
| sobrepes | 27.208 | 1.384 |
| obesitat | 33.241 | 3.327 |

Tal i com ja passava en l'altra mostra, el grup amb una variabilitat més alta és el dels obesos. Per contra, el que té una menor dispersió és el grup de l'infrapès. Si testem aquestes diferències entre variàncies (tal i com ja ho hem fet en l'apartat 4.1.2.), torna a sortir significativa amb un p.valor inferior a $2,2 \cdot 10^{-16}$.

Si comparem les mitjanes internes en aquest cas amb les que hem obtingut per la mostra del 1995, veiem que pels dos grups corresponents a adipositats baixes presenten mitjanes similars. En canvi, els grups del sobrepès i la obesitat presenten unes mitjanes lleugerament superiors en la mostra del 2006.

- **Edat:**

Val la pena donar un cop d'ull a la distribució de l'edat de la mostra del 2006 per extreure'n la màxima informació possible. A més, en el nostre cas és especialment interessant donat que en posteriors anàlisis es crearà una categorització per l'edat i s'estratificaran els resultats per aquesta.

| Min. | 1st Qu. | Median | Mean | St.Dev | 3rd Qu. | Max. | NA's | n |
|-------|---------|--------|-------|--------|---------|--------|---------|-------|
| 16.00 | 35.00 | 47.00 | 49.02 | 17.92 | 63.00 | 104.00 | 3273.00 | 29476 |

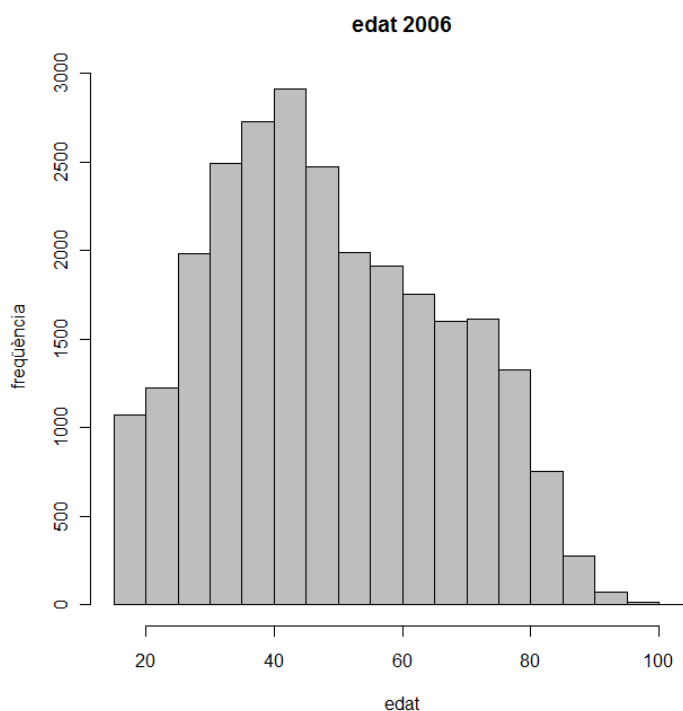


Figura 4.15. Histograma on s'hi representa l'edat de la mostra del 2006.

S'observa clarament una censura inferior, causada pel fet de treballar amb una mostra formada únicament per individus adults que per definició no poden tenir menys de 16 anys (tal i com els cataloga l'INE).

A més d'això, l'histograma corresponent a l'edat (4.15.) presenta una forma força estranya apuntant a una lleugera bimodalitat. Aquesta seria preocupant i caldria actuar-hi (estratificant per alguna variable com el sexe), si no fos perquè concorda força amb la piràmide poblacional corresponent a l'any anterior (figura 4.16).



Figura 4.16. Piràmide poblacional espanyola de l'any 2005. Font: INE, cens de l'1 de gener del 2005.

- **Sexe:**

Una última variable, sempre rellevant i també en el nostre cas donat que la usarem per a estratificar els nostres anàlisis, és el sexe. Així que creiem que val molt la pena obtenir-ne informació.

| dona | home |
|-------|-------|
| 15309 | 10894 |

Pel que fa al 2006, la mostra presenta més dones que homes, tot i que la diferència és del 58,4% respecte el 41,6%. No és una diferencia molt exagerada, però veiem que en aquesta mostra les proporcions s'assimilen més

a la realitat de la població que no pas en la mostra del 1995 (com hem pogut observar en l'apartat 4.1.2).

4.2.3. Descriptiva bivariant

En aquest apartat es pretén veure si hi ha algunes relacions entre les variables presentades anteriorment.

- **Alçada VS pes:**

En primer lloc es vol veure si hi ha relació entre l'alçada i el pes dels individus de la mostra del 2006. És força intuïtiu creure que poden presentar alguna relació, i ja hem vist que en la mostra del 1995 es donava, així que donarem un cop d'ull a aquestes dues variables conjuntament.

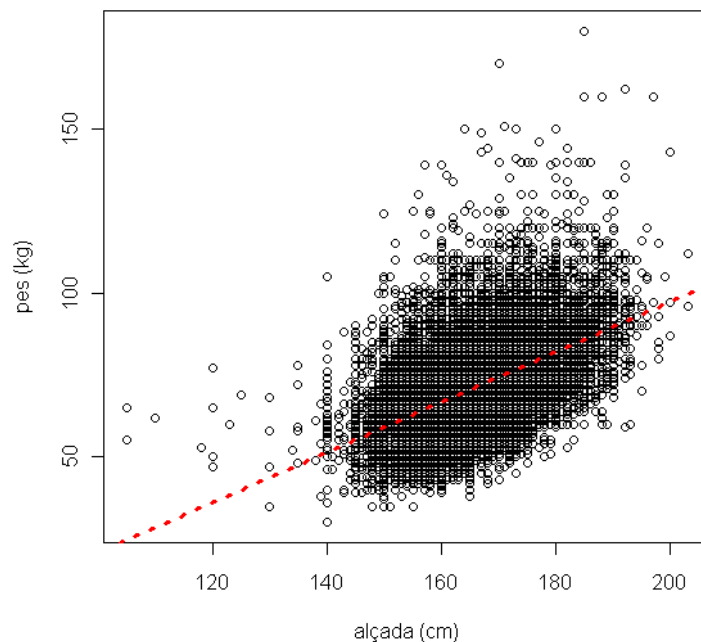


Figura 4.17. Diagrama de punts on es representen els valors dels individus per l'alçada i el pes en el 2006.

Veient el diagrama de punts presentat en la figura 4.17., s'observa una relació significativa entre aquestes dues variables; indicant-nos que individus amb valors grans d'alçada tendeixen cap a pesos alts.

Aquesta relació és força natural donada la complexió de les persones, i tot plegat ja lliga amb els resultats obtinguts en el 1995 i amb els coneixements previs. Cal destacar, però, que la recta que s'hi ajusta explica poc les dades.

- **Edat VS IMC:**

Com ja hem fet per l'altra mostra, donem un cop d'ull a aquesta relació amb la intenció d'intentar respondre si a mida que ens anem fent grans augmenta la nostra adipositat (en termes generals).

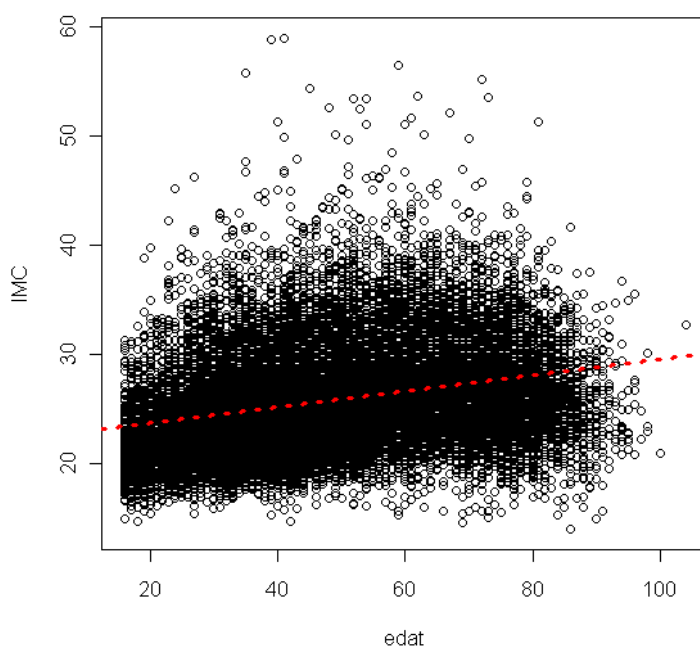


Figura 4.18. Diagrama de punts on es representen els valors dels individus per l'edat i l'IMC en el 2006.

Observant la figura 4.18. veiem que la relació en aquest cas és molt similar a la que es donava en la mostra del 1995. Així que, com en aquell cas, veiem que hi ha una relació dèbil però significativa entre ambdues variables.

De la mateixa manera que en la mostra anterior, això ens indica que quan anem creixent va augmentant la nostra adipositat; tot i no ser una relació que expliqui massa casos, i menys en aquesta ocasió en que tenim tants.

- **IMC estratificant per sexe:**

Donada la naturalesa del nostre estudi, ens ha semblat molt interessant crear aquesta variable categòrica a partir de l'IMC calculat anteriorment. Com que coneixem que, en termes generals, homes i dones presenten complexions diferents; ens sembla molt interessant veure les diferències entre sexes per l'IMC.

| sexe | n | Mean | St. Dev. | Min. | 1st Qu. | Median | 3rd Qu. | Max. |
|-------------|-------|--------|----------|-------|---------|--------|---------|-------|
| Dona | 15309 | 25.356 | 4.710 | 13.96 | 21.91 | 24.68 | 28.00 | 58.96 |
| Home | 10894 | 26.531 | 3.923 | 14.88 | 23.88 | 26.12 | 28.71 | 53.57 |

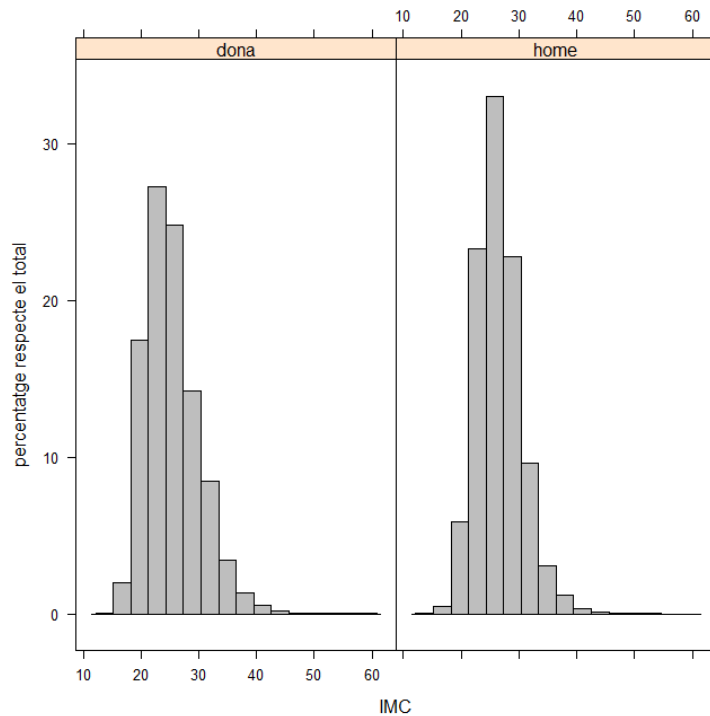


Figura 4.19. Histogrames de l'IMC del 2006 estratificant per sexe.

Numèricament destaquen les diferències entre les mitjanes i les desviacions tipus, tal i com passava amb la mostra del 1995. Observem que les dones presenten una mitjana per sota de la dels homes i que aquests darrers presenten una dispersió menor que les dones en quant a l'adipositat.

De la mateixa manera que abans testegem aquestes diferències mitjançant l'estadístic F de Fisher-Snedecor per la variància, i el de la t d'Student per les mitjanes (apartat 4.1.3.). En ambdós casos obtenim p.valors inferiors a $2,2 \cdot 10^{-16}$; fet que ens porta a rebutjar la hipòtesis d'igualtat de variàncies en el primer cas, i el d'igualtat de mitjanes en el segon.

La diferència en la dispersió prové del fet que les dones arriben a valors més baixos que els homes, i també valors més alts. El primer ja es donava en el cas de les dades del 1995, però el fet que les dones presentin un màxim superior al dels homes no es donava abans. Així que, d'una manera encara més exagerada en l'altra base de dades, això provoca que el rang de les dones sigui

major al dels homes; fet que concorda amb les diferències en les desviacions tipus.

De la figura 4.19. el més destacable és que, en aquest cas, també és molt més simètrica la distribució dels homes que la de les dones. Això concorda amb la descriptiva numèrica, que ens ho indica amb la similitud entre la mitjana i la mediana en el cas dels homes. Tot i així, les dues variables presenten unes cues destacables en la part alta de la distribució; si bé és cert que en el cas de les dones la cua és molt més pesada que en els dels homes.

En l'histograma també podem veure que la distribució dels homes és més estilitzada que la de les dones, que ja hem vist que presenten una variabilitat superior.

5. COMPARATIVA DE LES DISTRIBUCIONS DE L'IMC DEL 1995 I 2006

5.1. AJUST DE DISTRIBUCIONS

Tal i com s'ha exposat en la introducció, un dels principals objectius que ens hem marcat és comparar les distribucions de l'IMC del 2006 respecte el del 1995. Per fer-ho, sembla una bona idea buscar quina distribució de probabilitat coneguda que s'ajusti bé a cadascuna de les variables; i posteriorment dur a terme una comparació paramètrica entre ambdues (IMC del 1995 i el del 2006).

L'estratègia seguida ha sigut de prova i error, testant diferents distribucions mitjançant el test de Kolmogorov-Smirnov (exposat en l'apartat 4.1.2.). Com que en les corresponents descriptives univariants havíem vist que no procedien d'una distribució normal, s'han provat altres distribucions de probabilitat contínues que creiem que podrien ajustar-se havent observat la forma dels histogrames (weibull, lognormal i gamma). Per a cadascuna s'ha estimat prèviament els paràmetres més adequats per a la mostra del 1995 i la del 2006.

Per la weibull, el test ens porta a refusar la hipòtesi nul·la (els p.valors es presenten a la taula 5.1.); i a més a més al sobreposar la distribució teòrica a les dades, es veu clarament que no concorden.

| | IMC 1995 | IMC 2006 |
|------------------|-------------------------|------------------------|
| Weibull | $< 2,2 \cdot 10^{-16}$ | $< 2,2 \cdot 10^{-16}$ |
| Gamma | $< 2,118 \cdot 10^{-6}$ | $< 2,2 \cdot 10^{-16}$ |
| Lognormal | 0,002167 | $7,9 \cdot 10^{-9}$ |

Taula 5.1. Taula de p.valors corresponents al test de K-S per les variables i les distribucions indicades.

Al provar amb una distribució gamma, també refuseu H_0 ; però si veiem el gràfic, veiem que és aquesta la distribució que més bé s'ajusta de totes. I el cas que millor s'ajusta a les dades dels tres és la lognormal; que tampoc passa el test de bondat d'ajust, però amb un test més poder a l'acceptació i una representació gràfica que ens convenç força. També cal recordar que estem treballant amb tamanyos de mostra molt grans, fet que fa realment difícil que s'accepti la hipòtesis nul·la en algun test de bondat d'ajust.

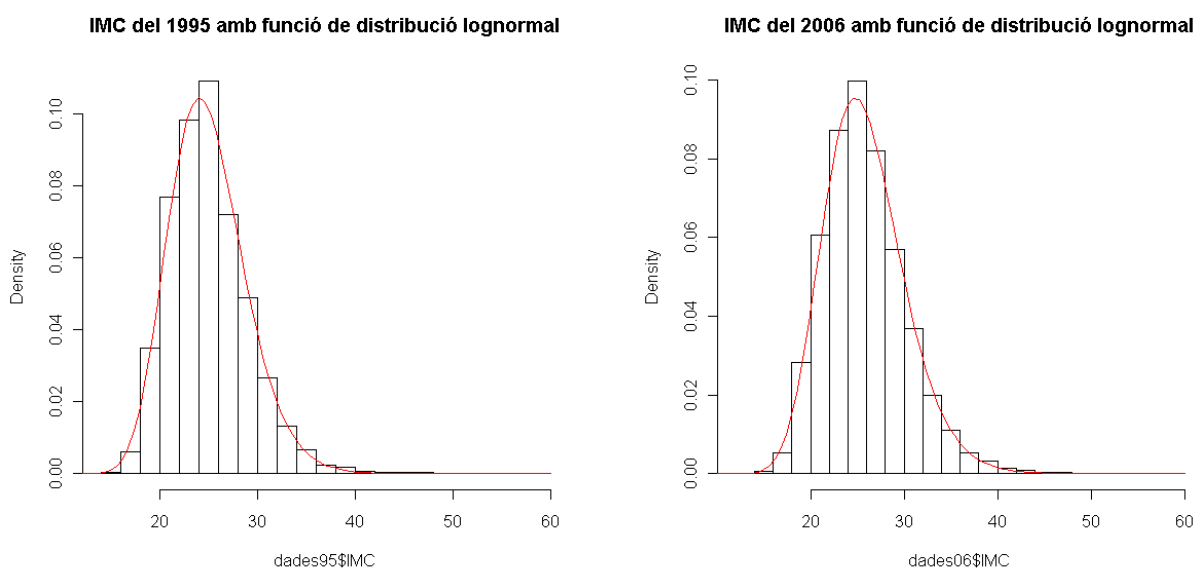


Figura 5.1. Distribucions dels IMC sobreposant-hi una distribució lognormal.

Així, que no hem pogut trobar cap funció de probabilitat que s'ajusti adequadament a les variables de l'IMC (per cap dels dos anys); i per tant es fa impossible dur a terme una comparació paramètrica.

Com que no hem pogut ajustar un model paramètric, farem una anàlisi no paramètrica; que no implica l'ajust de cap model en concret.

En primer lloc, realitzarem el test de K-S per a dues mostres; per tal de respondre a la pregunta de si les dues distribucions són iguals o no, és a dir,

$$\begin{cases} H_0: \text{Les dues variables provenen de la mateixa distribució.} \\ H_1: \text{Les dues variables no provenen de la mateixa distribució.} \end{cases}$$

Aquest test funciona exactament igual que el de K-S d'una sola mostra, però en aquest cas comparem dues funcions de distribució mostrals (en comptes de comparar una mostral amb una teòrica).

L'estadístic de prova, en aquest cas és el següent:

$$D_{n,n'} = \sup_x |F_{1,n}(x) - F_{2,n'}(x)|, \text{ on en aquest cas } F_{1,n} \text{ i } F_{2,n'}, \text{ són les}$$

distribucions empíriques de cadascuna de les variables.

Els resultats de passar aquest test a les variables corresponent a l'IMC del 1995 i el del 2006 es presenten a continuació, i en ells es pot veure clarament com refusem la hipòtesis nul·la. Això ens indica que no podem dir que les dues variables provinguin de la mateixa distribució, fet que ens indica que hi ha diferències entre la distribució de l'IMC de la mostra del 1995 i el del 2006.

Two-sample Kolmogorov-Smirnov test

data: dades95\$IMC and dades06\$IMC

D = 0.0904, p-value < 2.2e-16

alternative hypothesis: two-sided

Així doncs, s'arriba a la conclusió que entre el 1995 i el 2006 s'han produït canvis pel que fa a la distribució de l'IMC. Tot seguit procedirem a cercar la manera com avaluar aquests canvis, per poder fer-ho en els següents apartats.

5.2. FUNCIONS DE DENSITAT

Per tal de comparar les distribucions de l'IMC de les dues mostres, un dels objectius principals d'aquest treball; calcularem les seves corbes de densitat estimades. Ho farem mitjançant un procediment de tipus kernel, en el que

s'aniran calculant els punts corresponents a les corbes a partir d'un nucli gaussià; per després unir-los i formar aquestes funcions de densitat estimades.

La metodologia emprada és la mateixa pels dos anys, assegurant així que són el més comparables possible. Tot i això, cal tenir en compte que la corba corresponent al 2006 ha estat estimada amb un nombre major d'individus que la del 1995.

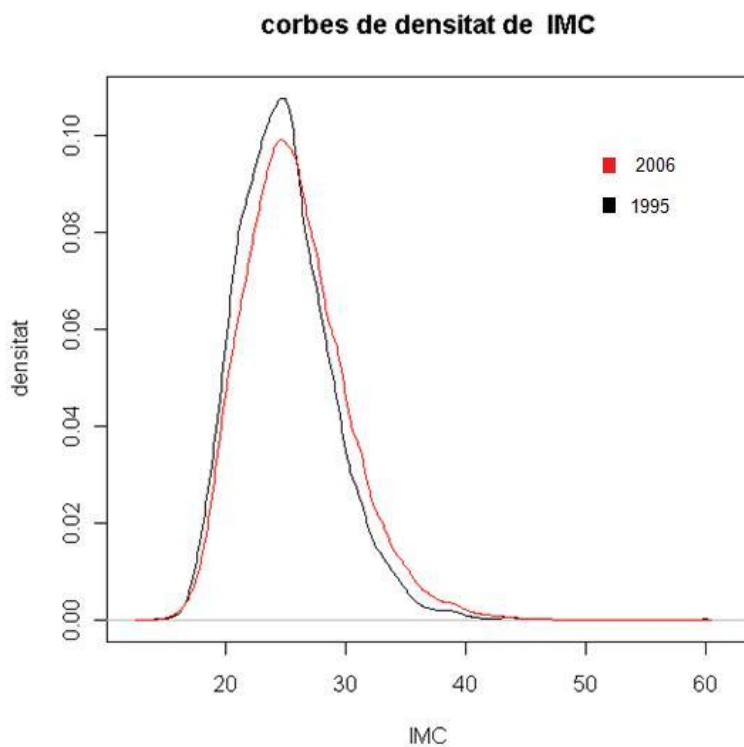


Figura 5.2. Representació de les corbes corresponents a les funcions de distribució estimades de l'IMC de l'any 2006 i del 1995.

Observant-les (figura 5.2.), el primer que crida l'atenció és que la distribució corresponent al 2006 presenta una cua notablement més llarga i pesada que la del 1995. Això fa que tingui una forma més aplanada i que, per tant, es tracti d'una distribució amb més variabilitat. Així ho corrobora la descriptiva numèrica presentada a continuació:

| | Min. | Median | Mean | St.Dev | Max. | NA's | n |
|----------------|-------|--------|-------|--------|-------|---------|-------|
| any1995 | 14.02 | 24.62 | 25.00 | 4.07 | 59.88 | 991.00 | 6389 |
| any2006 | 13.96 | 25.39 | 25.84 | 4.44 | 58.96 | 3273.00 | 29476 |

D'una manera gràfica, en la figura 5.2. es torna a destacar el que ja hem provat en l'apartat anterior (5.1.): les distribucions de l'IMC de l'any 1995 i del 2006 són diferents.

Si bé és cert que no hi ha un canvi exagerat, s'aprecien diferències en la forma i es pot arribar a intuir un lleuger desplaçament cap a la dreta de la totalitat de la campana. A més cal tenir en compte que d'una mostra a l'altra només han transcorregut 10 anys, temps en el qual tampoc s'espera observar uns canvis radicals en aquests aspectes.

El canvi en la forma, donat sobretot pel creixement de la cua superior, ens denota que en la mostra del 2006 hi ha una major tendència a IMC força elevats entre els individus amb adipositat alta; mentre que en la del 1995 els individus amb IMC alts no presentaven uns valors tant extrems. Pel que fa al desplaçament generalitzat de la corba, ens estaria indicant que hi ha una tendència general de la població a presentar adipositats majors; la qual cosa no vol dir que no apareguin valors més baixos, ja que donat l'augment de la variabilitat, la corba corresponent al 2006 presenta una amplada major que la del 1995 a arribar a valors més baixos.

Així que en termes generals s'han produït aquests dos canvis: la distribució del 2006 s'ha desplaçat lleugerament cap a la dreta, i presenta una variabilitat major produïda sobretot per una cua superior més pesada.

Havent vist aquests canvis, i coneixent que la complexió física dels homes i de les dones és força diferent, sembla una bona idea fer el mateix però separant els uns dels altres. D'aquesta manera podrem observar si els diferents sexes es comporten o no de manera diferent. Així aprofitarem també per assegurar-nos que les diferents proporcions d'aquests en les mostres no estigui alterant

les nostres observacions; ja que si homes i dones es comportessin de maneres molt diferents, ens hauríem d'assegurar dues mostres amb les mateixes proporcions d'uns i altres si el que volem és comparar-ho globalment.

Sigui com sigui, sempre és una bona idea estratificar per una variable com el sexe quan s'està treballant amb variables de l'estil del pes i l'alçada (entre altres).

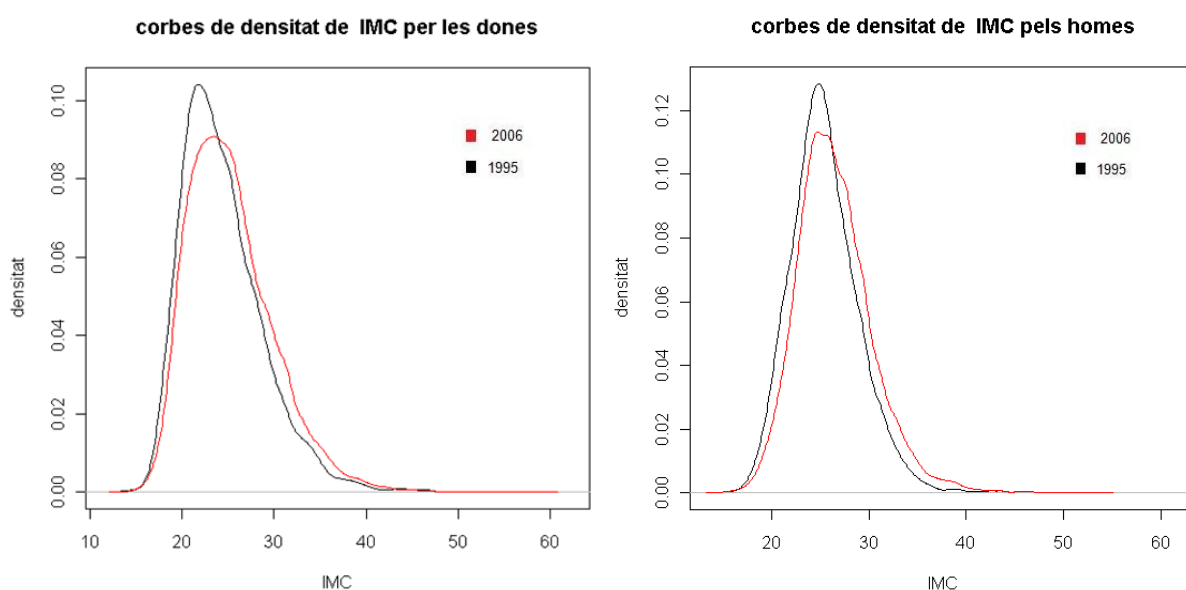


Figura 5.3. Representació de les corbes corresponents a les funcions de distribució estimades de l'IMC de l'any 2006 i del 1995 pels diferents sexes.

A la vista de la figura 5.3., sembla que el comportament general de les corbes no varia molt entre homes i dones. Si bé és cert que els homes es centren en un valor nominal major que les dones, i que ells es caracteritzen per tenir unes corbes més estilitzades que elles; la comparativa entre els diferents anys resulta similar en ambdós sexes.

Tant en el cas dels homes com en el de les dones, les diferències observades són les mateixes que en el cas global; tenint en compte els diferents comportaments per sexes comentats anteriorment. El que més sobte de la figura 5.3. és que, a l'estratificar per sexes, el desplaçament generalitzat de les

corbes s'accentua. Aquest fet és degut a que la mostra del 2006 presenta un percentatge major de dones (58,4%) que la del 1995 (48,2%); i al ser elles les que sempre tendeixen a presentar un IMC més baix, suavitzen el desplaçament en la representació global (figura 5.2.).

5.3. TUKEY MEAN-DIFFERENCE PLOT

Per tal de seguir analitzant els canvis en les distribucions de l'IMC, s'ha emprat el gràfic de mitjanes-diferències de Tukey (Tukey m-d plot).

Aquesta eina gràfica consisteix en dividir les distribucions que es desitja comparar en quantils, i posteriorment graficar les mitjanes i les diferències entre una distribució i l'altre. Es procedeix calculant els valors corresponents a cadascun dels quantils per a les dues distribucions, i a continuació es crea el gràfic situant la diferència entre els dos mateixos quantils de cadascuna de les distribucions, en l'eix d'ordenades (y); i la mitjana entre els dos valors dels quantils en l'eix d'abscisses (x).

Per exemple, si ens trobéssim en el cas que el quantil de 0.3 per la primera distribució fos 30 i el de la segona 25, en l'eix d'ordenades ens situaríem en el 5 (o -5, depenent de quina distribució s'agafi com a basal); i en el d'abscisses en el 27'5 (mitjana entre els dos quantils).

Per tant, si dues distribucions són idèntiques, els quantils serien iguals i les diferències totes zero. Quedant així tots els punts situats damunt d'una recta en el zero, que prendrem com a referència. És senzill copsar que quan comparem distribucions diferents (el que realment ens interessa); la situació dels punts del gràfic m-d de Tukey en l'eix d'ordenades, ens indicarà la direcció (per sobre o per sota de la recta $y=0+n$) i la magnitud del canvi (molta o poca distància respecte la recta $y=0+n$).

Com que és molt més senzill d'entendre-ho a base d'exemples, tot seguit en presentem alguns on es pot observar com respon el gràfic m-d de Tukey a diferents situacions.

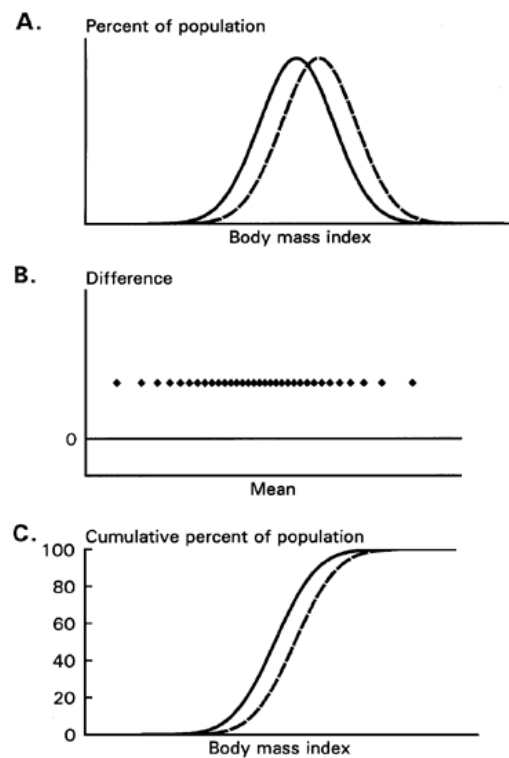


Figura 5.4. Primer exemple teòric per veure el comportament del Tukey m-d plot.

Font: annex 9.5.

En la primera (figura 5.4.), ens trobem amb dues distribucions amb la mateixa forma però on la segona distribució està desplaçada cap a la dreta. Aquest canvi és recollit pel gràfic de mitjanes i diferències, presentant un salt en l'eix d'ordenades respecte la recta de referència.

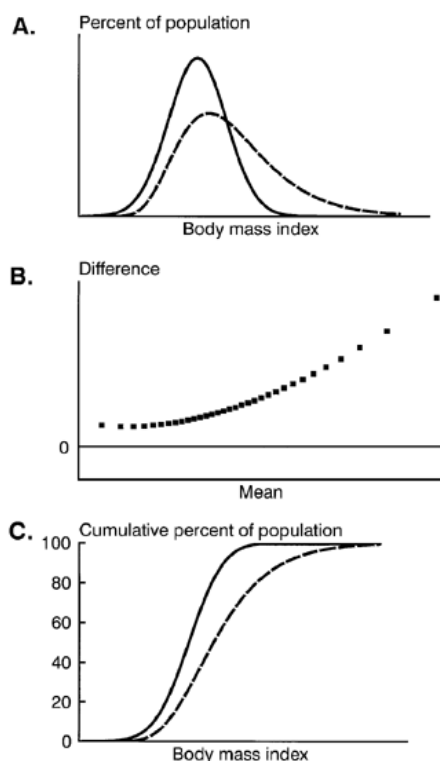


Figura 5.5. Segon exemple teòric de l'explicació del Tukey mean difference plot. Font: annex 9.5.

Pel que fa al segon exemple (figura 5.5.), es tracta d'un cas en el que la segona distribució s'assembla força a la primera per a valors baixos però presenta una cua pesada amb biaix cap a la dreta (valors alts). En aquest cas el gràfic m-d de Tukey també atrapa molt bé la informació, ja que en els valors baixos tenim un petit salt respecte a la recta de referència, i a mesura que augmenten els valors s'observa aquest increment de les diferències.

Es fa evident que aquesta és una eina clara i pràctica per comparar distribucions. Però cal tenir en compte que els exemples presentats es donarien en cas de distribucions teòriques, i que per tant, nosaltres mai ens trobarem en situacions tan clares i marcades. Un exemple del que ens podríem trobar treballant amb distribucions mostrals (com és el nostre cas), és el que es presenta a la figura 5.6., en la que podem veure que, tot i que les dues distribucions són molt semblants, es presenta una oscil·lació degut al soroll i un creixement a les cues (ja que al haver-hi pocs individus, petits canvis aquí es faran molt exagerats).

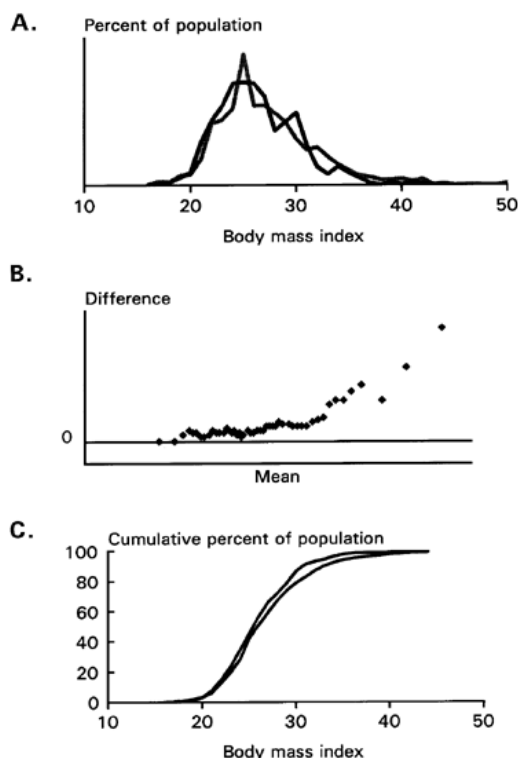


Figura 5.6. Exemple del Tukey m-d plot en un possible cas mostral. Font: annex 9.5.

Un cop exposat i comprés el funcionament d'aquesta eina gràfica, procedim a mostrar els resultats obtinguts al haver-la utilitzat per efectuar la comparativa que ens ocupa. Cada "Tukey mean-difference plot" (Tukey m-d plot) es mostra al costat del corresponent gràfic resultant de la representació conjunta de les dues corbes de densitat a comparar. D'aquesta manera pretenem tenir una visió més completa del que està passant, i permetre una millor comprensió del comportament del Tukey m-d plot i dels canvis que s'hi representen.

En tots els gràfics de mitjanes-diferències de Tukey presentats a continuació, s'han calculat les diferències restant els valors corresponents al 1995 dels del 2006, tal i com ens sembla més natural ("Període Actual" – "Període Anterior").

Per elaborar els gràfics de Tukey, no s'han considerat els individus amb valors majors a 40 per l'IMC; ja que es tractava d'un percentatge molt baix de la població (menys d'un 0,4% del 1995 i menys d'un 0,7% del 2006). El que hem fet ha estat tractar aquestes dades com anomalies, evitant així el risc de

calcular mitjanes amb un únic individu i de mostrar molta variabilitat en les cues del gràfics.

Per començar, es presenta a la figura 5.7. la comparativa del global de tota la població del 1995 amb la del 2006; tal i com s'havia efectuat en la figura 5.2. però afegint el Tukey m-d plot.

En ella observem clarament com hi ha hagut un desplaçament de la mitjana a causa dels canvis en el pes de la cua superior. També hi destaquen unes diferències negatives en la part esquerra de la distribució, provocades per la presència de valors més baixos del 2006 causats per la major amplada d'aquesta distribució.

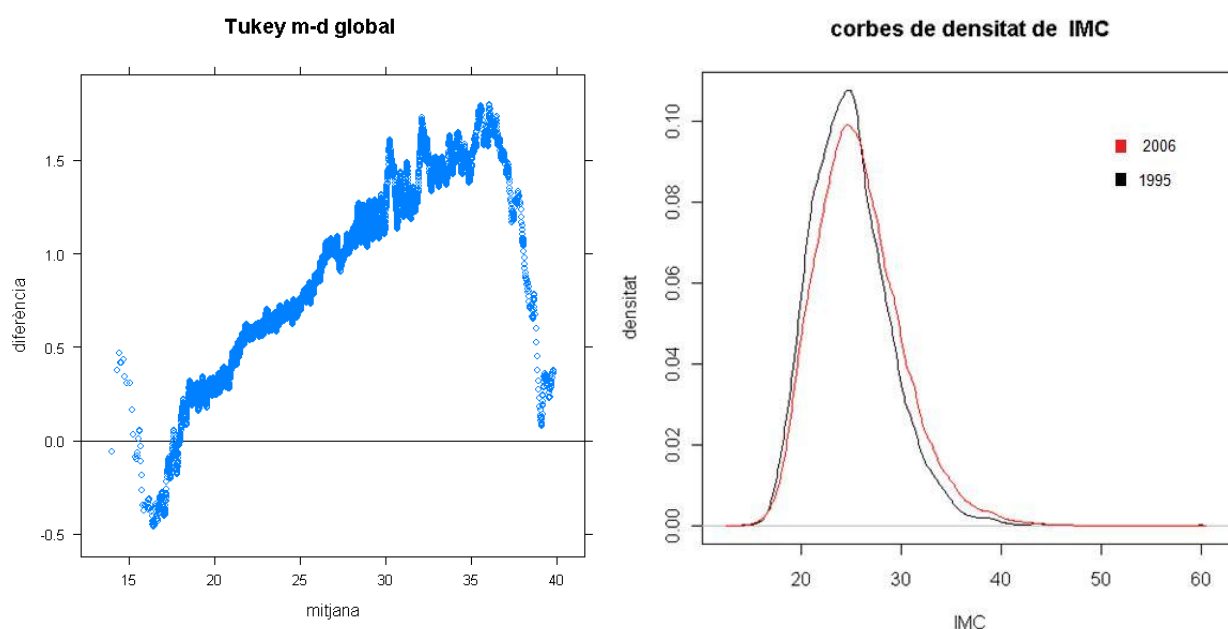


Figura 5.7. Tukey m-d plot i corbes de densitat comparant les dades del 2006 amb les del 1995 globalment.

Pel comportament del Tukey m-d plot, podem veure com ens indica un desplaçament de la distribució. Aquest canvi apunta a que la població presenta adipositats majors; així que torna a aparèixer aquesta tendència a que la gent cada cop sigui més grassa. Tot i tractar-se d'un petit canvi (l'escala així ens ho indica), cal recordar que estem treballant amb una distància d'aproximadament 10 anys; així que ja és d'esperar no trobar uns canvis exagerats.

En la comparativa global, com en tots els Tukey m-d plot's presentats a continuació, s'observen punts amb comportaments semi-erràtics en les parts extremes de les distribucions. Això és degut a que, en aquestes zones, hi ha pocs individus; la qual cosa ens du a presentar-hi punts molt afectats pel soroll. Tot i ser conscients d'això, els hem mantingut en els nostres càlculs.

Mitjançant la visió global de les mostres concloem que, com ja havíem pogut intuir en l'apartat 5.2., entre la distribució de l'IMC del 1995 i la del 2006 hi ha un desplaçament. El Tukey m-d plot també ens indica un canvi entre les distribucions, fet que concorda amb el que ja hem vist anteriorment: la corba corresponent al 2006 presenta una cua superior força més pesada i llarga que la del 1995.

Com s'ha vist en l'apartat 5.1., és convenient efectuar la comparativa separant els individus mitjançant el seu sexe. No és d'estranyar que sigui així donades les diferències en la complexió d'homes i dones. Per tant, també efectuarem aquesta estratificació per elaborar el Tukey m-d plot; i a continuació (figues 5.8. i 5.9.) se'n mostren els resultats.

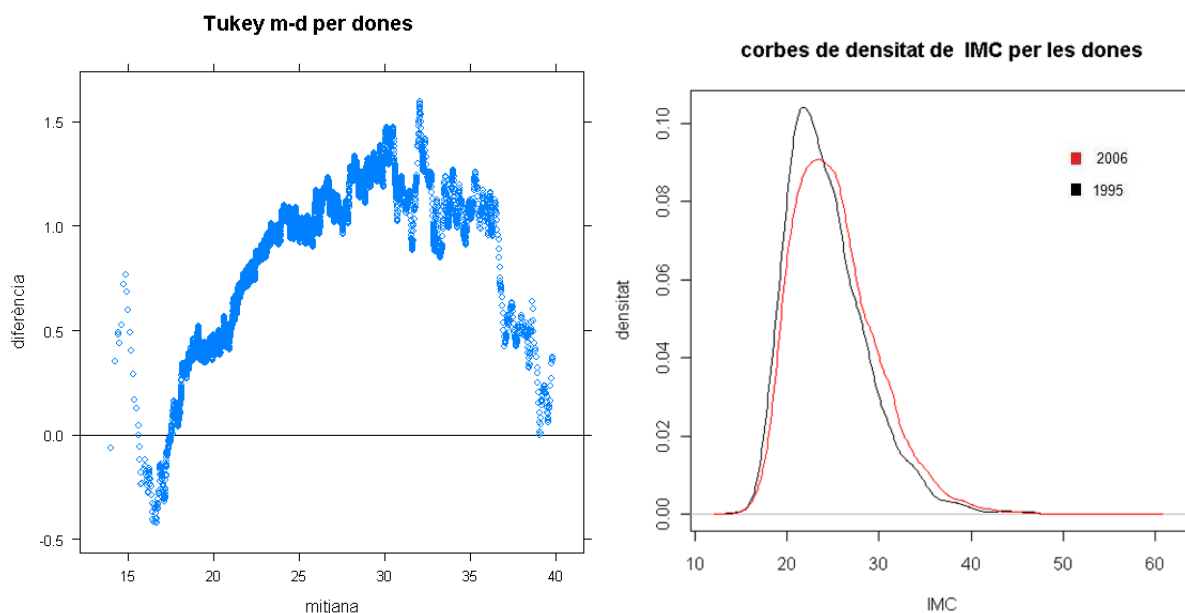


Figura 5.8. Tukey m-d plot i corbes de densitat per les dones. Comparant les dades del 2006 amb les del 1995.

Pel que fa a les dones (figura 5.8.), s'observa el canvi en el comportament pel que fa a valors de l'IMC elevats. Tal i com veiem en l'observació global, la forma de la corba de densitat ha canviat i així ho indica el gràfic de Tukey; denotant un lleuger desplaçament cap a la seva part alta respecte l'ordenada. Però un cop més, cal tenir en compte que trobem punts sobre la recta del zero i fins i tot per sota d'ella. Com ja hem comentat, això és degut a la major variabilitat en la distribució corresponent al 2006.

Vist tot això, sembla que en aquest cas s'intueix una mica més la tendència al desplaçament generalitzat de la corba; respecte a la visió global de les mostres. De totes maneres no és que sigui exagerat, i com abans, el que més destaca és el canvi en la forma de la corba de densitat; amb aquest augment del pes de la cua corresponent als valors alts de l'IMC.

Si passem a donar un cop d'ull al comportament dels homes (figura 5.9.), veiem que difereix una mica respecte al de les dones.

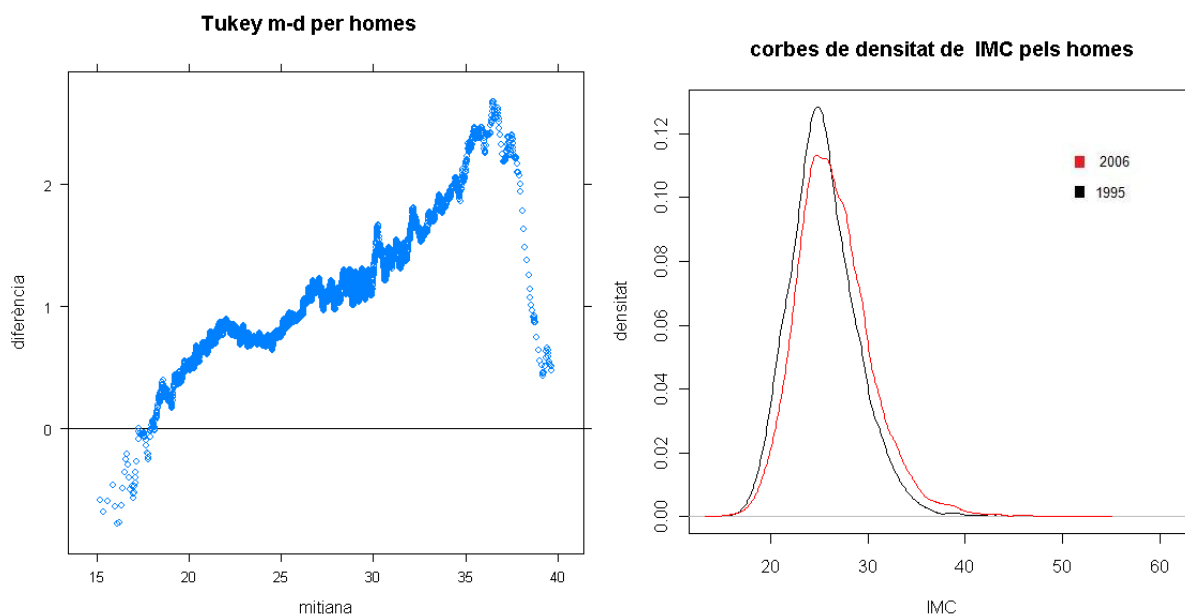


Figura 5.9. Tukey m-d plot i corbes de densitat pels homes comparant les dades del 2006 amb les del 1995 .

Efectuant la comparativa únicament amb les corbes de densitat, ja copsàvem que les corbes dels homes són més estretes i altes que les de les dones (en

ambdós anys). Si bé és així, havíem observat canvis similars pels dos sexes; i això no acaba de ser cert. En el cas dels homes es presenten canvis una mica més notables, com ens destaca l'escala del gràfic de Tukey. A més, en aquest cas el desplaçament generalitzat de la corba de densitat sembla més clar.

També s'observa el canvi de forma, com ens apunta aquesta corba notable del Tukey m-d plot; i torna a ser degut a la cua pesada dels valors alts de l'IMC, com indica el fet de trobar unes diferències majors a valors alts de la mitjana.

Vistes i analitzades les diferències estratificant per sexe, ens sembla interessant fer una anàlisi estratificant també per grups d'edat, per tal d'identificar on es produeixen els canvis observats en les distribucions globals. A més a més, aquesta estratificació ens permetrà comparar els resultats a Espanya amb els obtinguts en els EE.UU; mostrats en l'article "Changes in distribution of body mass index in the US population" (annex 9.5.) publicat per l'"International Journal of Obesity". D'aquest article és d'on s'han agafat els intervals per la categorització de l'edat. La diferència entre l'article esmentat i la nostra anàlisi, és que la nostra població adulta comença als 16 anys (tal i com ho fixa l'INE); així que hem decidit agrupar aquest individus amb els grup dels joves. També cal destacar que en el grup dels ancians, s'ha agrupat la gent de més de 70 anys; però cal recordar que com ja s'ha comentat anteriorment, la nostra comparativa només arriba fins als 85 anys.

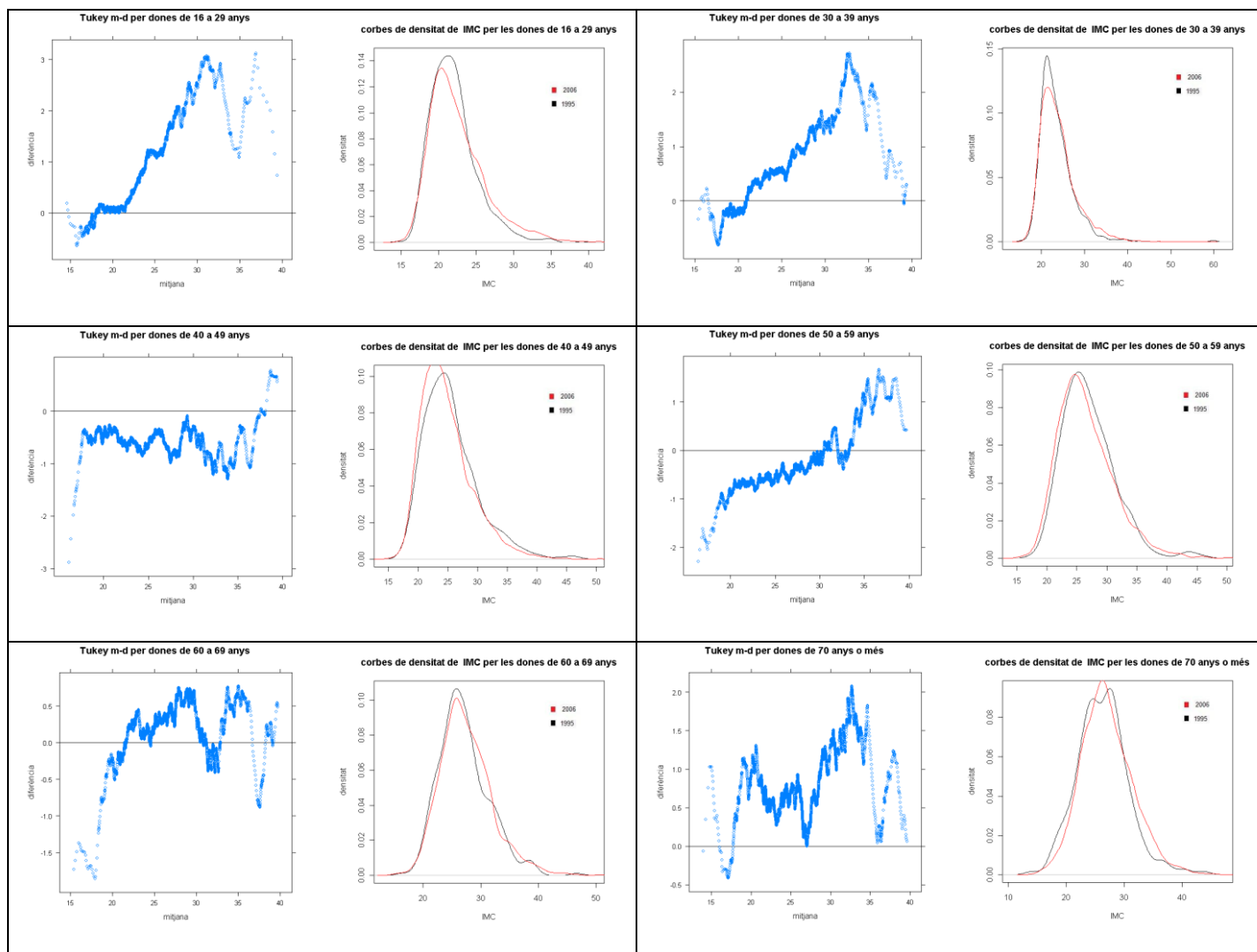


Figura 5.10. Tukey m-d plot i corbes de densitat per les dones estratificant pels diferents grups d'edat comparant les dades del 2006 amb les del 1995.

Observant les dones separades en grups d'edat (figura 5.10.), podem veure que presenten diferents comportaments. Així com en alguns grups destaquen canvis en la part esquerra del gràfic (corresponent a persones amb un IMC baix), en altres edats no trobem diferències en aquesta zona.

Es pot veure que els grups de la gent jove (entre els 16 i els 39 anys) es comporten de manera similar. Així que podríem dir que, en aquestes edats, les dones amb adipositat baixa es comporten de manera similar en els dos anys. Per aquest col·lectiu no sembla que hi hagi canvis en la part de les corbes de densitat que els pertoca (la corresponent a adipositats baixes), ja que apareixen solapades i properes al zero en el gràfic de Tukey. Per contra, si en

aquests mateixos grups d'edat observem les dones amb adipositat notable o alta; veiem que aquí sí que hi ha una tendència a augmentar aquesta adipositat, sobretot en el col·lectiu més jove. Així que, pel que fa a aquests grups d'edat, es destaca que hi ha un canvi en la forma, donat que les persones grasses ho són d'una manera més exagerada en la mostra del 2006. Per tant, el seu comportament és similar al del global en aquest aspecte.

Si passem a observar altres grups d'edat (dins la mateixa figura), podem veure comportaments realment sorprenents. Pel que fa a la nostra mostra, veiem que les dones entre 40 i 49 anys presenten un comportament totalment diferent a la resta. En aquest cas s'apunta a un desplaçament general de la corba, però cap a estar cada vegada més primes. El desplaçament és petit (com ens mostra l'escala del gràfic de Tukey), però sembla prou generalitzat com per destacar-lo tractant-se d'una diferència de 10 anys entre les mostres.

En el grup de les dones de 50 a 59 anys la tendència que s'apunta és que les dones amb adipositat baixa tendeixen a estar cada cop més primes, i que les dones grasses tendeixen a valors d'IMC majors. Això provoca un augment de la variabilitat, i així ens ho indica el gràfic de Tukey mostrant lleugeres diferències negatives per les mitjanes baixes i petites diferències positives per les mitjanes altes. També es pot veure un aplanament proper a la recta del zero per la zona central del gràfic, que ens estaria indicant que les persones amb valors proper a l'IMC ideal i una mica superiors ($IMC=25$ i una mica més) no han canviat gaire el seu comportament.

Pel que fa als grups de dones en una edat més avançada, no es podia considerar que hi hagi canvis destacables; ja que les corbes apareixen força solapades i el Tukey m-d plot respon presentant comportament erràtics propers a la recta del zero (donats pel soroll).

Si procedim a fer la mateixa estratificació en els homes (figura 5.11.), observem que aquests es comporten d'una manera força diferent a les dones. En el seu cas sembla que sí que es pot trobar un comportament generalitzat del canvi, sigui quin sigui el grup d'edat; i aquest comportament coincideix lògicament amb el general.

Deixant de banda el grup d'homes entre 60 i 69 anys, en general els homes tendeixen a canviar la forma de la corba de la densitat en la seva part alta. Si bé és cert que en alguns grups també es nota lleugers desplaçaments generalitzats per sobre la recta del zero (sobretot en el grup del homes entre 50 i 59 anys, i el dels homes més vells); bàsicament copsem que els homes amb valors baixos corresponents a l'IMC, segueixen presentant adipositats similars. Per contra, els homes amb adipositat alta, presenten una tendència molt marcada i generalitzada cap a l'augment d'aquesta.

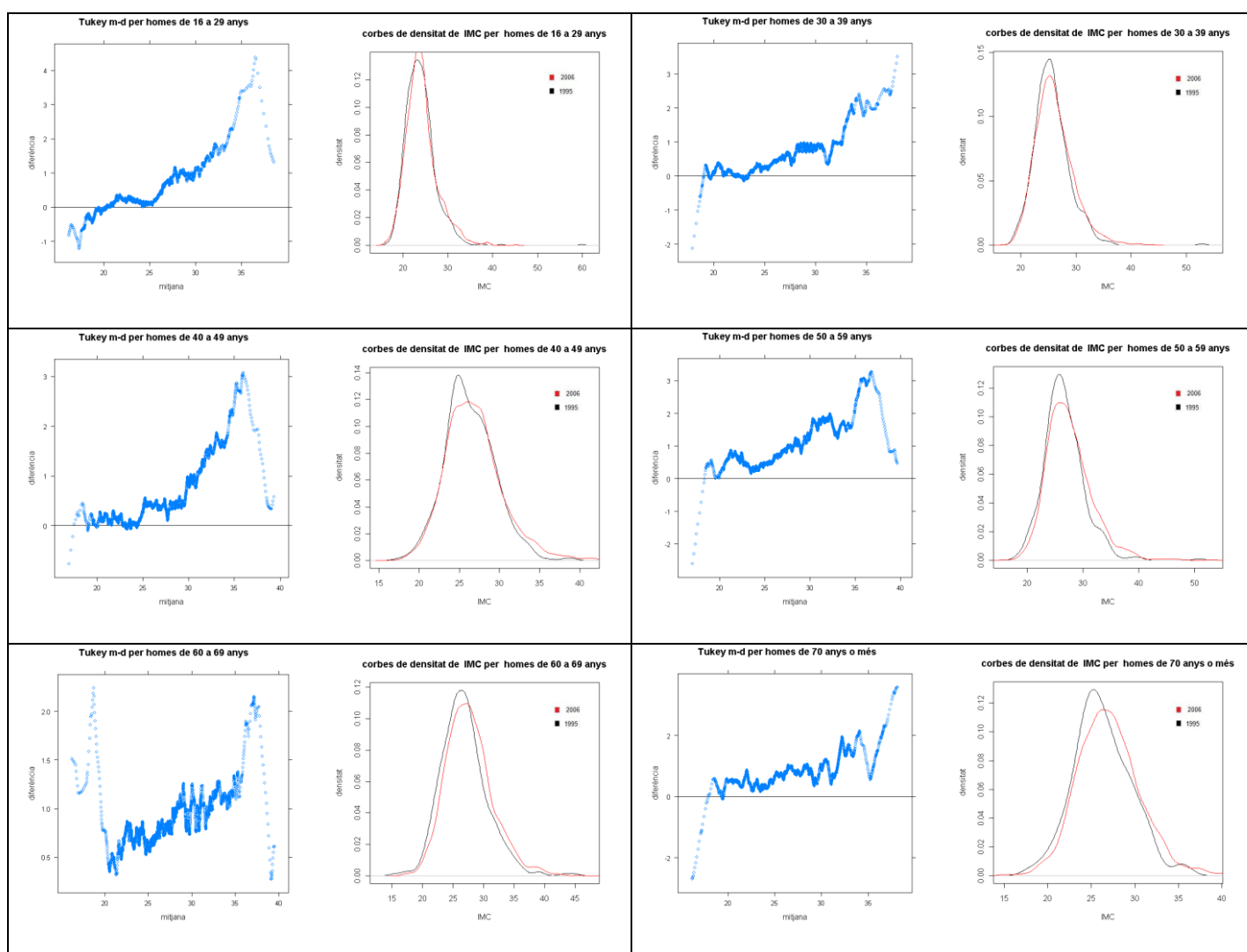


Figura 5.11. Tukey m-d plot i corbes de densitat pels homes estratificant pels diferents grups d'edat. Comparant les dades del 2006 amb les del 1995.

El grup d'edat que es comporta de forma molt diferent a la resta, és el dels homes entre 60 i 69 anys. Aquest col·lectiu presenta un desplaçament generalitzat i notable (donada la poca separació temporal de les mostres) de tota la corba de densitat. És tan general el desplaçament, que en el gràfic de Tukey la recta del zero apareix com a eix d'abscisses, destacant que no hi ha cap valor de la mitjana en el que s'hagi produït una disminució de l'adipositat del 2006 respecte la del 1995.

Quelcom similar passa en el grup del homes entre 59 i 60 anys i els més ancians (majors de 70 anys), que com ja hem comentat s'apropen més al comportament general dels homes; però també denoten un desplaçament general de les corbes de densitat, donat que es troben pocs individus per sota de la recta del zero o sobre d'ella.

Així doncs, si comparem els homes i les dones estratificant-los pels grups d'edat (figures 5.10., i 5.11.), el més destacable és que els homes presenten més una tendència en conjunt; on els grassos cada cop són més grassos i on en notables ocasions hi ha un desplaçament generalitzar de la corba de densitat. Per contra, les dones es comporten de maneres molt diferents en els diferents grups d'edat; fent així impossible marcar un patró de comportament en elles. I si parlem de conclusions generals podríem dir que hi ha un petit canvi i que tendim a estar més grassos (sobretot entre els grups amb adipositat alta).

5.4. COMPARATIVA ENTRE ESPANYA I EE.UU.

Per comparar els nostres gràfics amb els que es presenten en l'article americà ja esmentat (annex 9.5.), el primer que cal tenir en compte és que les nostres bases de dades han estat recollides amb 11 anys de diferència i les americanes amb 14. No és una diferència exagerada, i nosaltres ja vam procurar que fos així; però de totes maneres s'ha de tenir en compte perquè és una diferència existent.

També cal considerar que l'estudi dels Estats Units compara una mostra dels anys 80 amb una dels 90, i en el nostre cas estem comparat una mostra del 1995 amb una del 2006. No sabem si això pot comportar diferències donat que es tracta d'èpoques diferents, per això cal tenir-ho present. I val la pena apuntar que nosaltres només arribem a valors de 40 corresponents a l'IMC, perquè com ja s'ha comentat teníem molts pocs individus per sobre i els hem considerat anòmals; en canvi ells no han acotat i arriben a valors de 45.

Volem apuntar també que, els dos estudis s'han elaborat mitjançant les enquestes nacionals de salut corresponents, fet que els fa força comparables donat que les estratègies de mostreig i la mida de les mostres són similars.

Un cop situats, procedim a elaborar la comparativa: observant els gràfics nord-americans (annex 9.5.) copsem que en tots apareix una tendència molt similar entre homes i dones per diferents grups d'edat. En general, es veu com la distribució és molt semblant en la part corresponent a valors baixos de l'IMC; i després es produeix un augment de les diferències quan la mitjana va augmentant. És cert que per alguns grups d'edat totes les diferències apareixen per sobre de la recta del zero, sobretot en el cas de les dones; indicant un desplaçament generalitzat de la distribució de l'IMC. Però el que sobretot ens crida l'atenció, és que en l'estudi dels EE.UU. el comportament és molt semblant en tots els casos, cosa que no hem vist en la nostra anàlisi; on justament s'ha observat que per diferents sexes i diferents grups d'edat hi ha comportaments molt diferents pel que fa al canvi de les distribucions.

També destaca que les diferències als Estats Units són més exagerades que les nostres (ells arriben a diferències de l'ordre de 8 i nosaltres no superem diferències de 4); però abans de treure conclusions errònies cal veure que les majors diferències es donen en els valors més alts de l'IMC (com en la majoria dels nostres casos), i nosaltres no arribem a valors tan alts de l'IMC com l'estudi dels americans.

De totes maneres, sí que és cert que sembla que hi ha augments més destacats en els EE.UU. que en el nostre cas. Això ens indica que els canvis en les distribucions de l'IMC entre els anys 80 i els 90 als EE.UU., són majors que els que s'han produït recentment en Espanya. També podem destacar que l'evolució nord-americana és molt més homogènia que la nostra per sexes i per grups d'edat; i que ells en tots els casos apunten a l'augment de l'adipositat (ja sigui generalitzat o només entre les persones que presenten valors alts de IMC), cosa que en les nostres dades no es dona.

Per tant, concloem la comparativa afirmant que el comportament de les nostres mostres no es pot equiparar a la de les de l'estudi americà (annex 9.5.); i que això pot ser causa d'una diferència poblacional entre els països, una diferència de comportament provocada per la manca de contemporaneïtat dels estudis, o una unió d'ambdues coses.

6. CORBES CORRESPONENTS ALS QUARTILS DE L'IMC

Un dels objectius d'aquest projecte fou aconseguir una eina gràfica similar a la presentada en la figura 3.3. per la població adulta, que ens permeti situar el nostre IMC respecte als percentils poblacionals estimats donant-nos una idea de com és d'estrany el nostre IMC en comparació amb la resta de la població. A més, vistos els gràfics de la figura 3.3., desconeixíem com continuava l'evolució de la nostra adipositat i teníem curiositat per saber que passava. No sabíem si les corbes dels adults no tenien interès perquè l'evolució s'estancava i continuava de forma lineal i plana; però dubtàvem que fos així. Per tot això vam creure força interessant l'elaboració de les corbes presentades en aquest capítol.

Els gràfics s'han elaborat amb la mostra de l'ENSE del 2006, per ser que és la més actual de les quals se'n posseeixen les microdades. Una possibilitat hagués estat la d'ajuntar ambdues mostres tractades (la del 1995 i la del 2006) aconseguint així un major nombre d'individus, i per tant uns percentils més acurats; però s'ha desestimat aquesta opció donat que el comportament respecte l'IMC ha canviat d'una mostra a l'altra, com hem pogut copsar en el capítol 5 d'aquest projecte.

També cal concretar que per causes de manca d'individus, els càlculs dels percentils que relacionen l'edat amb l'IMC, només arriben fins als 85 anys; ja que a partir d'aquesta edat apareixen pocs casos (al voltant del 15% dels individus), i a més es troben repartits en moltes edats diferents. Així que, s'ha decidit arribar només fins els 85 anys per evitar la falta de representativitat de les poques persones que es donen per cadascuna de les edats compreses entre aquesta i els 104 anys.

Dit tot això, també comentar que en els gràfics de les corbes apareixen pintades les zones corresponents a les diferents categories de l'IMC; seguint el

mateix codi de colors emprat en la taula 3.3., amb l'objectiu de facilitar la lectura d'aquesta eina gràfica als usuaris.

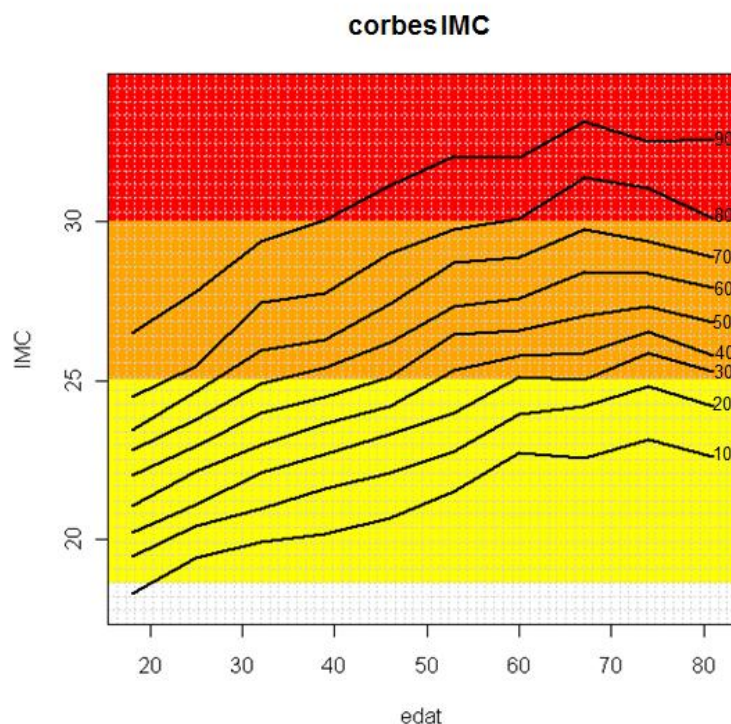


Figura 6.1. Corbes formades pels percentils corresponent a la mostra del 2006.

En la figura 6.1. presentem aquestes corbes pel global de la població sense separar per sexes. A més de per situar-se individualment, aquests gràfics també ens seran útils per veure el comportament de la mostra davant de l'IMC mitjançant una eina gràfica més.

Aquest primer gràfic no ens servirà tant per situar a individus concrets i comparar-los amb la resta; ja que com hem vist en apartats anteriors, els homes i les dones presentem comportaments força diferents respecte a l'adipositat. Així que per comparar-nos, seran més adequats els gràfics estratificant per sexes (figures 6.2. i 6.3.).

Un cop vist això, de la figura 6.1. en podem extreure informacions visuals generals. El primer que copsem és que major part de les corbes es troben

situades sobre les zones del normopès i del sobrepès (de color groc i taronja, respectivament). Això concorda amb el que s'ha vist en la descriptiva univariant del apartat 4.2.2.. També concorda el fet que hi ha molt pocs individus situats en la franja corresponent al infrapès (color blanc); i a més podem veure mitjançant les corbes que la major part d'aquest es troben en la zona de la gent més jove (entre els 16 i els 22 anys).

El que també veiem és que hi ha una tendència a l'augment de l'adipositat en l'inici, i un aplanament de les corbes al entrar en edats avançades. Així que, tot i que havíem vist que l'edat i l'IMC mantenen una relació significativa; estem còpsant que tot indica que aquesta relació no presenta una forma lineal.

En el gràfic global (figura 6.1.), també copsem que l'amplada de les corbes es manté força estable en les diferents edats. Aquest fet ens indica que no s'apunta a un canvi en la variabilitat de l'IMC per diferents edats, com a mínim no si ho mirem de manera general.

A continuació procedim a presentar les corbes estratificant per sexe. Com ja hem comentat aquestes són les més interessants si el que es pretén és situar un individu concret i comparar-lo amb la resta; sobretot tenint en compte les diferències existents entre les distribucions de l'IMC pels diferents sexes. Per començar explicarem com cal procedir per utilitzar de manera adient aquesta eina gràfica:

El primer que cal fer és conèixer l'edat, el pes i l'alçada de l'individu a situar en el gràfic. Mitjançant la fórmula presentada en l'apartat 3.2., calculem l'IMC de l'individu a partir de l'alçada i el pes; i un cop el tenim busquem el punt que correspon a la intersecció entre aquest valor i el de la seva edat. Al situar el punt en el gràfic corresponent al sexe de l'individu, podrem saber en quina categoria de l'IMC es troba. A més, també podem saber quin percentil és el més proper a ell; coneixent així quin percentatge de la població presenta unes adipositats superiors a l'individu tractat, i quins inferior.

Per exemple, una dona de 33 anys que presentés un IMC de 20,29; podria utilitzar la figura 6.2. (corresponent al seu sexe) per veure que es troba en la

franja del normopès. A més, sabria també que es troba entre el percentil del 20% i el del 30%; coneixent així que aproximadament el 25% de la població femenina de la seva edat presenta una adipositat inferior a la seva i el 75% superior.

Un cop conegut el funcionament de l'eina gràfica per comparar un individu concret amb la resta, podem donar un cop d'ull als gràfics de les corbes pels diferents sexes (figures 6.2. i 6.3.). El primer que cal destacar en ambdós casos, és la concordança en la continuïtat d'aquestes amb les famoses corbes de creixement dels infants (figura 3.3.). Si es miren els quantils representats tant per als adults com per als nens, es pot veure que les coincidències són destacables; indicant-nos una concordança important per a la fiabilitat de les nostres corbes.

Comparant les corbes dels diferents sexes, observem que les dels homes estan situades a una alçada major que les de les dones. Algunes de les repercussions d'aquest fenomen és que, aquell col·lectiu de joves que presentaven infrapès, veiem que es troben gairebé tots en el gràfic corresponent a les dones. Cal recordar que cadascuna de les corbes presentades és un percentil (valor de qual s'indica en els gràfics al costat de la corba). Per aquesta raó, encara que no hi hagi cap corba damunt de la zona de l'infrapès en el cas dels homes, no podem afirmar que no hi hagi cap individu en aquella zona; donat que no coneixem on es situen concretament els individus que conformen el 10% de la població que es troba per sota del primer percentil representat. Això succeeix en els tres gràfics amb aquesta naturalesa (figures 6.1., 6.2. i 6.3.), així que cal anar amb molt de compte amb la seva interpretació.

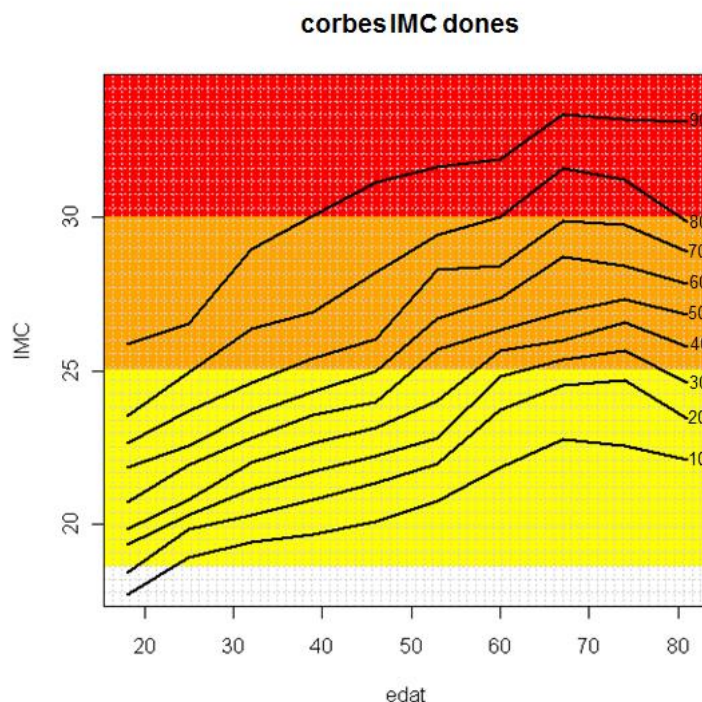


Figura 6.2. Corbes formades pels percentils corresponent a les dones de la mostra del 2006.

Si seguim comparant els dos sexes, podem notar com les corbes corresponents a les dones són més amples; fet que ens indica que són elles les que presenten una major variabilitat en l'adipositat, concordant un cop més amb el que s'intuïa en l'anàlisi univariant de l'apartat 4.2.2 i en el capítol 5. Així que sembla que els homes es comporten d'una manera més homogènia pel que fa a l'adipositat; però mitjançant aquests gràfics (figures 6.2. i 6.3.) no només copsem això, sinó que aquesta homogeneïtat es manté amb l'augment de l'edat.

Cal destacar també, que les corbes de les dones semblen aplanar-se més al arribar a edats avançades. De fet, fins hi tot es pot observar com presenten una lleugera tendència a la baixa, fet que ens indica que les dones tendeixen a aprimar-se a l'arribar a la vellesa.

En el cas dels homes d'edats més avançades cal destacar que les corbes s'ajunten més, la qual cosa ens indica que la variabilitat en l'adipositat tendeix a disminuir entre els homes ancians.



Figura 6.3. Corbes formades pels percentils corresponent als homes de la mostra del 2006.

7. CONCLUSIONS

El més destacable de la nostra anàlisi, és el fet d'haver observat que s'ha produït un canvi en les distribucions de l'IMC entre la mostra del 1995 i la del 2006. Ho hem comprovat mitjançant tests (apartat 5.1.), i de manera gràfica (en els apartats 5.2. i 5.3.). En totes les anàlisis coincidim en que les corbes de densitat de l'IMC no es comporten igual en ambdós anys; i el que s'observa a grans trets és una cua més pesada corresponent a adipositat altes, i un lleuger desplaçament del global de la distribució cap al sobrepès. Així que la resposta a la pregunta que ens formulàvem, és que en aquests darrers anys hi ha hagut una tendència a ser una població cada cop més grassa.

Si bé el que ens indiquen les dades tractades és aquesta tendència a l'augment generalitzat de l'adipositat; cal donar un cop d'ull més a fons, aprofitant al màxim la informació que posseïm. Tal i com hem vist en les corresponents descriptives bivariants (apartats 4.1.3. i 4.2.3.), en ambdues mostres s'observa una relació estadísticament significativa entre l'edat i l'adipositat (mesurada mitjançant l'IMC). Aquesta lleugera relació, ens indica que a mesura que anem creixent, tendim a presentar valors majors d'adipositat. Aquest fet, unit a les diferències entre les distribucions de l'edat entre la mostra del 1995 i la del 2006 (observables en els apartats 4.1.2. i 4.2.2.); podia estar explicant part (o la totalitat) d'aquest canvi en les distribucions de l'IMC.

Es pot comparar els histogrames corresponents a les edats (figures 4.6. i 4.15.), i veure que en el cas de la mostra del 1995 hi ha una destacable quantitat d'individus joves (entre els 16 i els 30 anys); on en la mostra del 2006 apareix un "forat" considerable. Aquest fenomen, recolzat per les respectives piràmides poblacionals (figures 4.7. i 4.16.), està mostrant el ja conegut envelliment de la població.

Donat aquest envelliment i la relació positiva entre adipositat i edat, pot ser que els canvis en les distribucions de l'IMC vinguin donats pels canvis en les edats dels individus de la nostra població; si més no, és evident que està relacionat.

Amb això no volem dir que la totalitat del desplaçament en les corbes de l'IMC, i l'augment en la cua superior vingui donada per aquest fenomen. Però estem convençuts que hi està afectant d'alguna manera (ja sigui decisiva o no), donada la relació observada i estadísticament demostrada entre les variables. A partir d'aquí ja els tocaria interpretar i cercar causes als especialistes.

L'altre focus d'atenció d'aquest projecte eren les corbes del quantils de l'IMC. Aquestes ens han indicat un fort posicionament de la major part dels individus (tant homes com dones) en els grups del normopès i el sobrepès. A més a més, en ha permès elaborar una comparativa entre ambdós sexes d'una manera molt visual; i ens serà útil per comparar-nos a nosaltres mateixos amb el comportament general estimat.

També cal destacar, d'aquestes corbes, el fet de que encaixen amb les que utilitzen els pediatres. Això, unit al gran nombre de dades emprades per la seva elaboració, ens fa creure en una elevada fiabilitat de les nostres corbes; quelcom sempre positiu.

I al dubte que ens plantejàvem de quina era la continuïtat d'aquelles corbes que serveixen per controlar el creixement dels nens, doncs ja hem pogut donar una resposta: les corbes encara triguen força a estancar-se (no ho fan fins entrada la vellesa). Així que, passada l'etapa de creixement, els individus encara tendeixen a presentar un augment de l'IMC; si bé és cert que amb un canvi notable de la pendent, com ja era d'esperar donada la natural i forta pujada que presenten les corbes del nens.

8. BIBLIOGRAFÍA

1. D.G. Altman, J.M. Bland, Measurement in Medicine: the Analysis of Method Comparison Studies. Any 1983.
2. M. J. Crawles, THE R BOOK. Ed: Wiley, edició de l'any 2007.
3. P. Dalgaard, Introductory Statistics with R, Ed: Springer, edició de l'any 2002
4. K.M. Flegal, R.P. Troiano, Changes in the distribution of body mass index of adults and children in the US population. International Journal of Obesity. Any 2000.
5. www.cis.es (Centro de Investigaciones Sociológicas).
6. www.ine.es (Instituto Nacional de Estadística).
7. www.msc.es (Ministerio de Sanidad, Política Social e Igualdad).
8. www.perseo.aesan.msps.es (Programa PERSEO).
9. www.wikipedia.org

9. ANNEX

9.1. NOTÍCIA SOBRE LES ENSE

EL PAÍS.COM | Sociedad

Inicio Internacional Política España Deportes Economía Tecnología Cultura Gente y TV Sociedad Opinión Blogs In English buscar

Educación | Salud | Ciencia | El Viajero | El País semanal | Domingo

ELPAÍS.com > Sociedad

Sanidad pone todos sus datos estadísticos al alcance de los usuarios

Mortalidad, morbilidad, dependencia y financiación se pueden consultar desde un portal

E. DE B. - Madrid - 04/10/2010

Vota ☆☆☆☆☆ Resultado ★★★★★ 20 votos

16 9

La dirección es <http://www.msc.es/estadEstudios/estadisticas/sisInfSanSNS/home.htm>, y tiene todos los datos estadísticos que posee el Ministerio de Sanidad del sistema de salud español. Financiación, mortalidad, estancias hospitalarias, medicamentos, gasto y opinión de los usuarios están entre los datos que se pueden consultar.

Como ejemplo de la utilidad de estos datos, la [Federación de Asociaciones para la Defensa de la Salud Pública](#) ha hecho ya una primera prospección, con estos resultados:

1.- Entre 1990 y 2010, el gasto sanitario público ha crecido (tomando como referencia el PIB) un 9%; el privado, el 37%.

2.- Disminuye la mortalidad por muchas de las patologías prevalentes, pero como contrapartida existe un incremento importante de la discapacidad. La esperanza de vida con buena salud se encuentra en España muy por debajo de la media de la UE-27.

3.- Existe una atención sanitaria cada vez más centrada en el hospital, con altas hospitalarias por procesos muy por encima de la media europea, con un retraso en la implantación de las alternativas a la hospitalización.

Ministerio de Sanidad y Asuntos sociales

A FONDO
Sede: Madrid (España)
Directivo: Jiménez, Trinidad(Ministra)
[Ver cobertura completa](#)

La noticia en otros webs

- webs en español
- en otros idiomas

Font: http://www.elpais.com/articulo/sociedad/Sanidad/pone/todos/datos/estadisticos/alcance/usuarios/elpepusoc/20101004elpepusoc_8/Tes

9.2. SCRIPT DE L'R

```
#lectura de dades de l'any 1995
infoarxiu95<-read.csv(file.choose(),sep=";")
infoarxiu95$LONGITUD<-as.numeric(infoarxiu95$LONGITUD)
dades95<-
read.fwf(file.choose(),col.names=infoarxiu95$VARIABLE,widths=infoarxiu95$LONGITUD)

#lectura de dades de l'any 2006
infoarxiu06<-read.csv(file.choose(),sep=";")
infoarxiu06$LONGITUD<-as.numeric(infoarxiu06$LONGITUD)
dades06<-
read.fwf(file.choose(),col.names=infoarxiu06$VARIABLE,widths=infoarxiu06$LONGITUD)

#a les dades del 95 se li han canviat els noms previament
#canviar noms a les variables del data.frame del 2006
colnames(dades06)[554] <- "pes"
colnames(dades06)[555] <- "alçada"
colnames(dades06)[11] <- "edat"
colnames(dades06)[12] <- "sexe"

#tractament dades mancants del 1995
dades95$pes[dades95$pes == 999] = NA
dades95$pes[dades95$pes == 998] = NA

dades95$alçada[dades95$alçada == 999] = NA
dades95$alçada[dades95$alçada == 998] = NA

#tractament dades mancants del 2006
dades06$pes[dades06$pes == 999] = NA
dades06$pes[dades06$pes == 998] = NA

dades06$alçada[dades06$alçada == 999] = NA
dades06$alçada[dades06$alçada == 998] = NA

#etiquetes del sexe (a més no estan codificats igual en els dos anys)
dades95$sexe[dades95$sexe==1]="home"
dades95$sexe[dades95$sexe==2]="dona"

dades06$sexe[dades06$sexe==1]="home"
dades06$sexe[dades06$sexe==6]="dona"
```

###DEPURACIÓ DE LES BASES DE DADES:

##1995:

#treure individu amb menys de 16 anys

dades95=dades95[with(dades95,dades95\$edat!=12),]

#treure individus (en aquest cas d'alçada menor a 80)

dades95=dades95[with(dades95,dades95\$alçada>80),]

#treure individus (en aquest cas el que pesa 695kg)

dades95=dades95[with(dades95,dades95\$pes!=695),]

##2006:

#treure individus (en aquest cas d'alçada menor a 80 però només és un que mesura 57)

dades06=dades06[with(dades06,dades06\$alçada>80),]

#consultar un individu(en aquest cas el del 2006 que pesa 25kg)

summary(dades06[with(dades06,dades06\$pes==25),])

#treure individus (en aquest cas el que te pes 25)

dades06=dades06[with(dades06,dades06\$pes!=25),]

#consultar un individu(en aquest cas el del 2006 que pesa 30kg)

summary(dades06[with(dades06,dades06\$pes==30),])

#crear variable IMC del 1995

dades95\$IMC<-((dades95\$pes)/(((dades95\$alçada)/100)^2))

#crear variable IMC del 2006

dades06\$IMC<-((dades06\$pes)/(((dades06\$alçada)/100)^2))

#crear IMC categoritzat del 1995

dades95\$IMCcat[dades95\$IMC<18.5]="infrapes"

dades95\$IMCcat[dades95\$IMC>=18.5&dades95\$IMC<25]="normopes"

dades95\$IMCcat[dades95\$IMC>=25&dades95\$IMC<30]="sobrepes"

dades95\$IMCcat[dades95\$IMC>=30]="obesitat"

#dades95\$IMCcat<-na.exclude(dades95\$IMCcat)

#crear IMC categoritzat del 2006

dades06\$IMCcat[dades06\$IMC<18.5]="infrapes"

dades06\$IMCcat[dades06\$IMC>=18.5&dades06\$IMC<25]="normopes"

dades06\$IMCcat[dades06\$IMC>=25&dades06\$IMC<30]="sobrepes"

```
dades06$IMCcat[dades06$IMC>=30]="obesitat"
#dades06$IMCcat<-na.omit(dades06$IMCcat)

#taules
table(dades95$IMCcat)
table(dades95$sexe)

table(dades06$IMCcat)
table(dades06$sexe)

#diagrames de sectors
pie(table(dades95$IMCcat),main="IMC categoritzat")

pie(table(dades06$IMCcat),main="IMC categoritzat")

#gràfics univariants
hist(dades95$alçada,main="alçada
1995",xlab="alçada(cm)",ylab="freqüència",col=8,breaks=20)
hist(dades95$pes,main="pes 1995",xlab="pes(kg)",ylab="freqüència",col=8,breaks=30)
hist(dades95$IMC,main="IMC 1995",xlab="IMC",ylab="freqüència",col=8,breaks=30)
hist(dades95$edat,main="edat 1995",xlab="edat",ylab="freqüència",col=8)

hist(dades06$alçada,main="alçada
2006",xlab="alçada(cm)",ylab="freqüència",col=8,breaks=20)
hist(dades06$pes,main="pes 2006",xlab="pes(kg)",ylab="freqüència",col=8,breaks=30)
hist(dades06$IMC,main="IMC 2006",xlab="IMC",ylab="freqüència",col=8,breaks=30)
hist(dades06$edat,main="edat 2006",xlab="edat",ylab="freqüència",col=8)

#test t-student
t.test(IMC~sexe,dades95)
t.test(IMC~sexe,dades06)

#test de Levene
library(NCStats)
leveneTest(IMC~sexe,dades95)
leveneTest(IMC~sexe,dades06)

leveneTest(IMC~IMCcat,dades95)
leveneTest(IMC~IMCcat,dades06)

#plots de normalitat
qqnorm(dades95$alçada,main="Normal Q-Q plot alçada",xlab="quantils
teòrics",ylab="quantils mostrals")
```

```
qqnorm(dades06$alçada,main="Normal Q-Q plot alçada",xlab="quantils  
teòrics",ylab="quantils mostrals")
```

```
##test de normalitat (kolmogorof-s)  
library(nortest)
```

```
lillie.test(dades95$alçada)  
lillie.test(dades95$pes)  
lillie.test(dades95$IMC)
```

```
lillie.test(dades06$alçada)  
lillie.test(dades06$pes)  
lillie.test(dades06$IMC)
```

```
#gràfics de punts bivariants del 1995  
plot(dades95$alçada,dades95$pes,xlab="alçada (cm)",ylab="pes (kg)")  
reg1=lm(dades95$pes~dades95$alçada)  
abline(reg1,col=2,lty=3,lwd=3)  
summary(reg1)
```

```
plot(dades95$edat,dades95$IMC,xlab="edat",ylab="IMC")  
reg2=lm(dades95$IMC~dades95$edat)  
abline(reg2,col=2,lty=3,lwd=3)  
summary(reg2)
```

```
#gràfics de punts bivariants del 2006  
plot(dades06$alçada,dades06$pes,xlab="alçada (cm)",ylab="pes (kg)")  
reg3=lm(dades06$pes~dades06$alçada)  
abline(reg3,col=2,lty=3,lwd=3)  
summary(reg3)
```

```
plot(dades06$edat,dades06$IMC,xlab="edat",ylab="IMC")  
reg4=lm(dades06$IMC~dades06$edat)  
abline(reg4,col=2,lty=3,lwd=3)  
summary(reg4)
```

```
#carregar llibreries (previament cal instal·lar-les)  
library(lattice)  
source("http://www.rforge.net/NCStats/InstallNCStats.R")  
library(NCStats)  
#grafiques estratificant per sexe  
histogram(~dades95$IMC|dades95$sexe,xlab="IMC",ylab="percentatge respecte el  
total",col=8)
```

```
histogram(~dades06$IMC|dades06$sexe,xlab="IMC",ylab="percentatge respecte el  
total",col=8)
```

```
#summary estratificant per sexe
```

```
Summarize(dades95$IMC~dades95$IMCcat,digits=3)
```

```
Summarize(dades06$IMC~dades06$IMCcat,digits=3)
```

```
Summarize(dades95$IMC~dades95$sexe,digits=3)
```

```
Summarize(dades06$IMC~dades06$sexe,digits=3)
```

```
#intentem ajustar distribució de probabilitat en IMC 1995 i IMC 2006
```

```
library(MASS)
```

```
#weibull
```

```
par=fitdistr(na.omit(dades06$IMC),"weibull")
```

```
ks.test(dades06$IMC, "pweibull", par$estimate[1],par$estimate[2])
```

```
hist(dades06$IMC,breaks=20,freq=F, main="IMC del 2006 amb funció de distribució weibull")
```

```
plot(function(x) dweibull(x,par$estimate[1],par$estimate[2]),xlim=c(0,60), add=TRUE,col=2)
```

```
par=fitdistr(na.omit(dades95$IMC),"weibull")
```

```
ks.test(dades95$IMC, "pweibull", par$estimate[1],par$estimate[2])
```

```
hist(dades95$IMC,breaks=20,freq=F, main="IMC del 1995 amb funció de distribució weibull")
```

```
plot(function(x) dweibull(x,par$estimate[1],par$estimate[2]),xlim=c(0,60), add=TRUE,col=2)
```

```
#lognormal
```

```
par=fitdistr(na.omit(dades06$IMC),"lognormal")
```

```
ks.test(dades06$IMC, "plnorm", par$estimate[1],par$estimate[2])
```

```
hist(dades06$IMC,breaks=20,freq=F, main="IMC del 2006 amb funció de distribució  
lognormal")
```

```
plot(function(x) dlnorm(x,par$estimate[1],par$estimate[2]),xlim=c(0,60), add=TRUE,col=2)
```

```
par=fitdistr(na.omit(dades95$IMC),"lognormal")
```

```
ks.test(dades95$IMC, "plnorm", par$estimate[1],par$estimate[2])
```

```
hist(dades95$IMC,breaks=20,freq=F, main="IMC del 1995 amb funció de distribució  
lognormal")
```

```
plot(function(x) dlnorm(x,par$estimate[1],par$estimate[2]),xlim=c(0,60), add=TRUE,col=2)
```

```
#gamma
```

```
par=fitdistr(na.omit(dades06$IMC),"gamma")
```

```
ks.test(dades06$IMC, "pgamma", par$estimate[1],par$estimate[2])
```

```
hist(dades06$IMC,breaks=20,freq=F, main="IMC del 2006 amb funció de distribució gamma")  
plot(function(x) dgamma(x,par$estimate[1],par$estimate[2]),xlim=c(0,60), add=TRUE,col=2)
```

```
par=fitdistr(na.omit(dades95$IMC),"gamma")
ks.test(dades95$IMC, "pgamma", par$estimate[1],par$estimate[2])
hist(dades95$IMC,breaks=20,freq=F, main="IMC del 1995 amb funció de distribució gamma")
plot(function(x) dgamma(x,par$estimate[1],par$estimate[2]),xlim=c(0,60), add=TRUE,col=2)

#fem el test per veure si IMC1995 i IMC2006 provenen de la mateixa distribució
ks.test(dades95$IMC,dades06$IMC)

#les dues densitats del IMC sobreposades
plot(density(na.omit(dades95$IMC)),main="corbes de densitat del
IMC",xlab="IMC",ylab="densitat")
lines(density(na.omit(dades06$IMC)),col=2)

#separem homes i dones
homes06=dades06[with(dades06,dades06$sexe=="home"),]
dones06=dades06[with(dades06,dades06$sexe=="dona"),]

homes95=dades95[with(dades95,dades95$sexe=="home"),]
dones95=dades95[with(dades95,dades95$sexe=="dona"),]

#densitats del IMC sobreposades separant per sexe
plot(density(na.omit(dones95$IMC)),main="corbes de densitat del IMC per
dones",xlab="IMC",ylab="densitat")
lines(density(na.omit(dones06$IMC)),col=2)
win.graph()
plot(density(na.omit(homes95$IMC)),main="corbes de densitat del IMC per
homes",xlab="IMC",ylab="densitat")
lines(density(na.omit(homes06$IMC)),col=2)

##test de normalitat (kolmogorof-s)
lillie.test(dades95$pes)

#creem categories per l'edat en les dues bases de dades
dades95$edatcat[dades95$edat<=29]="16-29"
dades95$edatcat[dades95$edat>29&dades95$edat<=39]="30-39"
dades95$edatcat[dades95$edat>39&dades95$edat<=49]="40-49"
dades95$edatcat[dades95$edat>49&dades95$edat<=59]="50-59"
dades95$edatcat[dades95$edat>59&dades95$edat<=69]="60-69"
dades95$edatcat[dades95$edat>=70]="70 o més"
dades95$edatcat=factor(dades95$edatcat,levels=c("70 o més","60-69","50-59","40-49","30-
39","16-29"))

dades06$edatcat[dades06$edat<=29]="16-29"
```

```
dades06$edatcat[dades06$edat>29&dades06$edat<=39]="30-39"  
dades06$edatcat[dades06$edat>39&dades06$edat<=49]="40-49"  
dades06$edatcat[dades06$edat>49&dades06$edat<=59]="50-59"  
dades06$edatcat[dades06$edat>59&dades06$edat<=69]="60-69"  
dades06$edatcat[dades06$edat>=70]="70 o més"  
dades06$edatcat=factor(dades06$edatcat,levels=c("70 o més","60-69","50-59","40-49","30-  
39","16-29"))
```

```
###per fer gràfic de mitjanes diferències de Tukey  
library(Rcmdr)  
dadesjunes <- mergeRows(dades06,dades95, common.only=TRUE)  
dadesjunes$any=c(rep("any06",times=nrow(dades06)),rep("any95",times=nrow(dades95)))  
dadesjunes$any=factor(dadesjunes$any,levels=c("any95","any06"))
```

```
#Tukey m-d  
#eliminem les persones amb IMC>40 (son menys del 0,7% en el 2006 i menys del 0,5% del  
1995)  
dadesjunes=dadesjunes[with(dadesjunes,dadesjunes$IMC<40),]
```

```
#tmd per tot el conjunt de dades amb les corbes de densitat  
win.graph()  
interm=qq(na.omit(dadesjunes$any~dadesjunes$IMC),panel =  
lattice.getOption("panel.qq"))#f.value=seq(0.01,0.99,0.000010))  
tmd(interm1,xlab="mitjana",ylab="diferència",main="Tukey m-d global")
```

```
plot(density(na.omit(dades95$IMC)),main="corbes de densitat del  
IMC",xlab="IMC",ylab="densitat")  
lines(density(na.omit(dades06$IMC)),col=2)
```

```
#separem homes i dones (en dadesjunes)  
dadesjunes$edatcat=factor(dadesjunes$edatcat,levels=c("50-59","60-69","70 o més","16-  
29","30-39","40-49"))  
juntesh=dadesjunes[with(dadesjunes,dadesjunes$sexe=="home"),]  
juntess=dadesjunes[with(dadesjunes,dadesjunes$sexe=="dona"),]
```

```
#fem els Tukey m-d per les dones  
win.graph()  
intermD=qq(na.omit(juntess$any~juntess$IMC),panel =  
lattice.getOption("panel.qq"))#f.value=seq(0.01,0.99,0.000010))  
tmd(intermD,xlab="mitjana",ylab="diferència",main="Tukey m-d per dones")
```

```
plot(density(na.omit(dones95$IMC)),main="corbes de densitat del IMC per les  
dones",xlab="IMC",ylab="densitat")  
lines(density(na.omit(dones06$IMC)),col=2)
```



```
#idem pels homes
win.graph()
intermH=qq(juntesh$any~juntesh$IMC,panel = lattice.getOption("panel.qq"))
tmd(intermH,xlab="mitjana",ylab="diferència",main="Tukey m-d per homes")

plot(density(na.omit(homes95$IMC)),main="corbes de densitat del IMC pels
homes",xlab="IMC",ylab="densitat")
lines(density(na.omit(homes06$IMC)),col=2)

#separem per categories d'edat dins dels diferents grups per sexes
#en les dones
jundesd20=jundesd[with(jundesd,jundesd$edatcat=="16-29"),]
jundesd30=jundesd[with(jundesd,jundesd$edatcat=="30-39"),]
jundesd40=jundesd[with(jundesd,jundesd$edatcat=="40-49"),]
jundesd50=jundesd[with(jundesd,jundesd$edatcat=="50-59"),]
jundesd60=jundesd[with(jundesd,jundesd$edatcat=="60-69"),]
jundesd70=jundesd[with(jundesd,jundesd$edatcat=="70 o més"),]

#i en els homes
juntesh20=juntesh[with(juntesh,juntesh$edatcat=="16-29"),]
juntesh30=juntesh[with(juntesh,juntesh$edatcat=="30-39"),]
juntesh40=juntesh[with(juntesh,juntesh$edatcat=="40-49"),]
juntesh50=juntesh[with(juntesh,juntesh$edatcat=="50-59"),]
juntesh60=juntesh[with(juntesh,juntesh$edatcat=="60-69"),]
juntesh70=juntesh[with(juntesh,juntesh$edatcat=="70 o més"),]

#separem per categories edat i sexe en les dades del 1995 i del 2006 per les corbes de densitat
d95d20=dones95[with(dones95,dones95$edatcat=="16-29"),]
d95d30=dones95[with(dones95,dones95$edatcat=="30-39"),]
d95d40=dones95[with(dones95,dones95$edatcat=="40-49"),]
d95d50=dones95[with(dones95,dones95$edatcat=="50-59"),]
d95d60=dones95[with(dones95,dones95$edatcat=="60-69"),]
d95d70=dones95[with(dones95,dones95$edatcat=="70 o més"),]

d95h20=homes95[with(homes95,homes95$edatcat=="16-29"),]
d95h30=homes95[with(homes95,homes95$edatcat=="30-39"),]
d95h40=homes95[with(homes95,homes95$edatcat=="40-49"),]
d95h50=homes95[with(homes95,homes95$edatcat=="50-59"),]
d95h60=homes95[with(homes95,homes95$edatcat=="60-69"),]
d95h70=homes95[with(homes95,homes95$edatcat=="70 o més"),]

d06d20=dones06[with(dones06,dones06$edatcat=="16-29"),]
```

```
d06d30=dones06[with(dones06,dones06$edatcat=="30-39"),]
d06d40=dones06[with(dones06,dones06$edatcat=="40-49"),]
d06d50=dones06[with(dones06,dones06$edatcat=="50-59"),]
d06d60=dones06[with(dones06,dones06$edatcat=="60-69"),]
d06d70=dones06[with(dones06,dones06$edatcat=="70 o més"),]

d06h20=homes06[with(homes06,homes06$edatcat=="16-29"),]
d06h30=homes06[with(homes06,homes06$edatcat=="30-39"),]
d06h40=homes06[with(homes06,homes06$edatcat=="40-49"),]
d06h50=homes06[with(homes06,homes06$edatcat=="50-59"),]
d06h60=homes06[with(homes06,homes06$edatcat=="60-69"),]
d06h70=homes06[with(homes06,homes06$edatcat=="70 o més"),]
#fem els tmd corresponents en el cas de les dones
win.graph()
intermD20=qq(na.omit(juntesd20$any~juntesd20$IMC),panel =
lattice.getOption("panel.qq"))#,f.value=seq(0.01,0.99,0.000010))
tmd(intermD20,xlab="mitjana",ylab="diferència",main="Tukey m-d per dones de 16 a 29
anys")
win.graph()
plot(density(na.omit(d95d20$IMC)),main="corbes de densitat del IMC per les dones de 16 a 29
anys",xlab="IMC",ylab="densitat")
lines(density(na.omit(d06d20$IMC)),col=2)

win.graph()
intermD30=qq(na.omit(juntesd30$any~juntesd30$IMC),panel =
lattice.getOption("panel.qq"))#,f.value=seq(0.01,0.99,0.000010))
tmd(intermD30,xlab="mitjana",ylab="diferència",main="Tukey m-d per dones de 30 a 39
anys")
win.graph()
plot(density(na.omit(d95d30$IMC)),main="corbes de densitat del IMC per les dones de 30 a 39
anys",xlab="IMC",ylab="densitat")
lines(density(na.omit(d06d30$IMC)),col=2)

win.graph()
intermD40=qq(na.omit(juntesd40$any~juntesd40$IMC),panel =
lattice.getOption("panel.qq"))#,f.value=seq(0.01,0.99,0.000010))
tmd(intermD40,xlab="mitjana",ylab="diferència",main="Tukey m-d per dones de 40 a 49
anys")
win.graph()
plot(density(na.omit(d95d40$IMC)),main="corbes de densitat del IMC per les dones de 40 a 49
anys",xlab="IMC",ylab="densitat")
lines(density(na.omit(d06d40$IMC)),col=2)

win.graph()
```

```
intermD50=qq(na.omit(juntesd50$any~juntesd50$IMC),panel =  
lattice.getOption("panel.qq"))#,f.value=seq(0.01,0.99,0.000010))  
tmd(intermD50,xlab="mitjana",ylab="diferència",main="Tukey m-d per dones de 50 a 59  
anys")  
win.graph()  
plot(density(na.omit(d95d50$IMC)),main="corbes de densitat del IMC per les dones de 50 a 59  
anys",xlab="IMC",ylab="densitat")  
lines(density(na.omit(d06d50$IMC)),col=2)  
  
win.graph()  
intermD60=qq(na.omit(juntesd60$any~juntesd60$IMC),panel =  
lattice.getOption("panel.qq"))#,f.value=seq(0.01,0.99,0.000010))  
tmd(intermD60,xlab="mitjana",ylab="diferència",main="Tukey m-d per dones de 60 a 69  
anys")  
win.graph()  
plot(density(na.omit(d95d60$IMC)),main="corbes de densitat del IMC per les dones de 60 a 69  
anys",xlab="IMC",ylab="densitat")  
lines(density(na.omit(d06d60$IMC)),col=2)  
  
win.graph()  
intermD70=qq(na.omit(juntesd70$any~juntesd70$IMC),panel =  
lattice.getOption("panel.qq"))#,f.value=seq(0.01,0.99,0.000010))  
tmd(intermD70,xlab="mitjana",ylab="diferència",main="Tukey m-d per dones de 70 anys o  
més")  
win.graph()  
plot(density(na.omit(d95d70$IMC)),main="corbes de densitat del IMC per les dones de 70  
anys o més",xlab="IMC",ylab="densitat")  
lines(density(na.omit(d06d70$IMC)),col=2)  
  
#i ara el mateix pels homes  
win.graph()  
intermH20=qq(na.omit(juntesh20$any~juntesh20$IMC),panel =  
lattice.getOption("panel.qq"))#,f.value=seq(0.01,0.99,0.000010))  
tmd(intermH20,xlab="mitjana",ylab="diferència",main="Tukey m-d per homes de 16 a 29  
anys")  
win.graph()  
plot(density(na.omit(d95h20$IMC)),main="corbes de densitat del IMC per homes de 16 a 29  
anys",xlab="IMC",ylab="densitat")  
lines(density(na.omit(d06h20$IMC)),col=2)  
  
win.graph()  
intermH30=qq(na.omit(juntesh30$any~juntesh30$IMC),panel =  
lattice.getOption("panel.qq"))#,f.value=seq(0.01,0.99,0.000010))
```

```
tmd(interH30,xlab="mitjana",ylab="diferència",main="Tukey m-d per homes de 30 a 39
anys")
win.graph()
plot(density(na.omit(d95h30$IMC)),main="corbes de densitat del IMC per homes de 30 a 39
anys",xlab="IMC",ylab="densitat")
lines(density(na.omit(d06h30$IMC)),col=2)

win.graph()
interH40=qq(na.omit(juntesh40$any~juntesh40$IMC),panel =
lattice.getOption("panel.qq"))#,f.value=seq(0.01,0.99,0.000010))
tmd(interH40,xlab="mitjana",ylab="diferència",main="Tukey m-d per homes de 40 a 49
anys")
win.graph()
plot(density(na.omit(d95h40$IMC)),main="corbes de densitat del IMC per homes de 40 a 49
anys",xlab="IMC",ylab="densitat")
lines(density(na.omit(d06h40$IMC)),col=2)

win.graph()
interH50=qq(na.omit(juntesh50$any~juntesh50$IMC),panel =
lattice.getOption("panel.qq"))#,f.value=seq(0.01,0.99,0.000010))
tmd(interH50,xlab="mitjana",ylab="diferència",main="Tukey m-d per homes de 50 a 59
anys")
win.graph()
plot(density(na.omit(d95h50$IMC)),main="corbes de densitat del IMC per homes de 50 a 59
anys",xlab="IMC",ylab="densitat")
lines(density(na.omit(d06h50$IMC)),col=2)

win.graph()
interH60=qq(na.omit(juntesh60$any~juntesh60$IMC),panel =
lattice.getOption("panel.qq"))#,f.value=seq(0.01,0.99,0.000010))
tmd(interH60,xlab="mitjana",ylab="diferència",main="Tukey m-d per homes de 60 a 69
anys")
win.graph()
plot(density(na.omit(d95h60$IMC)),main="corbes de densitat del IMC per homes de 60 a 69
anys",xlab="IMC",ylab="densitat")
lines(density(na.omit(d06h60$IMC)),col=2)

win.graph()
interH70=qq(na.omit(juntesh70$any~juntesh70$IMC),panel =
lattice.getOption("panel.qq"))#,f.value=seq(0.01,0.99,0.000010))
tmd(interH70,xlab="mitjana",ylab="diferència",main="Tukey m-d per homes de 70 anys o
més")
win.graph()
```

```
plot(density(na.omit(d95h70$IMC)),main="corbes de densitat del IMC per homes de 70 anys o  
més",xlab="IMC",ylab="densitat")  
lines(density(na.omit(d06h70$IMC)),col=2)
```

```
#corbes dels quantils de l'IMC (poblacio sencera)
```

```
A<-matrix(nrow=81,ncol=10)  
j=1  
for (i in seq(18,85,by=7)){  
  
  A[j,]=c(i,quantile(na.omit(dades06$IMC[dades06$edat==i]),probs=seq(0.1,0.9,0.1)))  
  j=j+1  
}
```

```
corves<-data.frame(A)  
#pintem les zones corresponents a la categoritzacio del IMC  
plot(0,type="n",xlim=c(0,1),ylim=c(0,1),axes="n",xlab="")  
rect(-1,0.04,2,0.44,col="yellow",border=F)  
rect(-1,0.44,2,0.75,col="orange",border=F)  
rect(-1,0.75,2,2,col="red",border=F)
```

```
##amb el "par" evitem que se'ns esborrin les zones de colors  
par(new=T)  
grid(70,40)  
par(new=T)  
matplot(corves$X1, corves[,  
c("X2","X3","X4","X5","X6","X7","X8","X9","X10")],col=rep(1,9),type="l",  
lty=1,lwd=2,xlab="edat",ylab="IMC", main="corbes IMC",ylim=c(18,34))
```

```
#separem homes i dones  
homes06=dades06[with(dades06,dades06$sexe=="home"),]  
dones06=dades06[with(dades06,dades06$sexe=="dona"),]
```

```
#corbes dels quantils de l'IMC per dones
```

```
A<-matrix(nrow=81,ncol=10)  
j=1  
for (i in seq(18,85,by=7)){  
  
  A[j,]=c(i,quantile(na.omit(dones06$IMC[dones06$edat==i]),probs=seq(0.1,0.9,0.1)))
```

```
j=j+1
}

corvesD<-data.frame(A)
#pintem les zones corresponents a la categorització del IMC
plot(0,type="n",xlim=c(0,1),ylim=c(0,1),axes="n",xlab="")
rect(-1,0.04,2,0.44,col="yellow",border=F)
rect(-1,0.44,2,0.75,col="orange",border=F)
rect(-1,0.75,2,2,col="red",border=F)

##amb el "par" evitem que se'ns esborrin les zones de colors
par(new=T)
grid(70,40)
par(new=T)
matplot(corvesD$X1, corvesD[, c("X2","X3","X4","X5","X6","X7","X8","X9","X10")],col=rep(1,9),
type="l", lty=1,lwd=2,xlab="edat",ylab="IMC",main="corbes IMC dones",ylim=c(18,34))

#corbes dels quantils de l'IMC per homes
win.graph()
A<-matrix(nrow=81,ncol=10)
j=1
for (i in seq(18,85,by=7)){

A[j,]=c(i,quantile(na.omit(homes06$IMC[homes06$edat==i]),probs=seq(0.1,0.9,0.1)))
j=j+1
}

corvesH<-data.frame(A)
#pintem les zones corresponents a la categorització del IMC
plot(0,type="n",xlim=c(0,1),ylim=c(0,1),axes="n",xlab="")
rect(-1,0.04,2,0.44,col="yellow",border=F)
rect(-1,0.44,2,0.75,col="orange",border=F)
rect(-1,0.75,2,2,col="red",border=F)

##amb el "par" evitem que se'ns esborrin les zones de colors
par(new=T)
grid(70,40)
par(new=T)
matplot(corvesH$X1, corvesH[, c("X2","X3","X4","X5","X6","X7","X8","X9","X10")],col=rep(1,9),
type="l", lty=1,lwd=2,xlab="edat",ylab="IMC",main="corbes IMC homes",ylim=c(18,34))
```

9.3. QUESTIONARI CORRESPONENT ALS ADULTS DE L'ENSE DE L'ANY 1995

Es pot trobar el qüestionari del 1995 complet a:

http://www.cis.es/cis/opencm/ES/1_encuestas/estudios/ver.jsp?estudio=2521&questionario=2552

A continuació es presenten les parts d'aquest qüestionari corresponents a les variables emprades en el nostre projecte.

A TODOS

P.32 ¿Podría decirme cuánto pesa, aproximadamente, sin zapatos ni ropa?. (Si está embarazada en la actualidad debe indicar cuánto pesaba antes del inicio del embarazo).

| | kilos | |
|-----------|-------|-----------------|
| N.S. | 998 | (282)(283)(284) |
| N.C. | 999 | |

P.33 ¿Y cuánto mide, aproximadamente, sin zapatos?

| | cms. | |
|-----------|------|-----------------|
| N.S. | 998 | (285)(286)(287) |
| N.C. | 999 | |

P.37 ¿Cuántos años cumplió Vd. en su último cumpleaños?

| | | |
|-----------|----|------------|
| N.C. | 99 | (293)(294) |
|-----------|----|------------|

A TODOS

P.40 Sexo:

| | | |
|---------------|---|-------|
| - Hombre | 1 | |
| - Mujer | 2 | (308) |

9.4. QUESTIONARI CORRESPONENT ALS ADULTS DE L'ENSE DE L'ANY 2006

Es pot trobar el qüestionari del 2006 complet a:
http://www.ine.es/metodologia/t15/ens_adu06.pdf

A continuació es presenten les parts d'aquest qüestionari corresponents a les variables emprades en el nostre projecte.

5.2 Edad

5.3 Sexo ☐ 1 Varón ☐ 6 Mujer

128. ¿Podría decirme cuánto pesa, aproximadamente, sin zapatos ni ropa?

Peso en Kg. →

No sabe/no recuerda ☐

No contesta ☐

129. ¿Y cuánto mide, aproximadamente, sin zapatos?

Altura en cm. →

No sabe/no recuerda ☐

No contesta ☐

9.5. ARTÍCULO NORD-AMERICÀ: “Changes in the distribution of body mass index of adults and children in the US population”

International Journal of Obesity (2000) 24, 807–818
© 2000 Macmillan Publishers Ltd. All rights reserved 0307-0565/00 \$15.00
www.nature.com/ijo



Changes in the distribution of body mass index of adults and children in the US population

KM Flegal¹* and RP Troiano^{1,2}

¹National Center for Health Statistics, Centers for Disease Control and Prevention, Hyattsville, MD, USA; and ²National Cancer Institute, NIH, Bethesda, MD, USA

BACKGROUND: National survey data show increases in mean body mass index (BMI) and in the prevalence of overweight and obesity for adults and children in the United States, indicating a change in the distribution of BMI.

OBJECTIVE: To apply graphical methods to describe changes in the distribution of BMI.

DESIGN: BMI values from the third National Health and Nutrition Examination Survey (NHANES III: 1988–94) were compared with data from earlier cross-sectional nationally representative surveys for adults 20–74 y of age and for children and adolescents 6–17 y of age. Tukey mean–difference plots were used to investigate the changes in the distributions of BMI within sex–age groups.

RESULTS: Mean–difference plots allow qualitative visual comparisons of the distributions of BMI between surveys. For all sex–age groups, there was increasing skewness with a greater shift in the upper part of the distribution so that, within each group, the heaviest subgroup was heavier in NHANES III than in prior surveys. For the youngest children, the lower part of the distribution showed virtually no change. With increasing age the whole distribution tended to shift upward slightly, suggesting an increase in BMI across the entire population.

CONCLUSIONS: These changes in the distribution of BMI suggest the combination of both profound environmental determinants and a population with a high degree of susceptibility. The reasons for the increasing prevalence of obesity should be sought in part by seeking to understand the factors causing increases in the population as a whole. *International Journal of Obesity* (2000) 24, 807–818

Keywords: adults; body weight; body mass index; children; health surveys; overweight; obesity; time trends; United States

Introduction

The prevalence of overweight and obesity have increased markedly in the United States for both children and adults.^{1–3} National survey data show that between 1976–80 and 1988–94 the age-adjusted prevalence of obesity (body mass index (BMI) ≥ 30.0) increased by 8 percentage points, from 14.5% to 22.5%, in the US adult population ages 20–74 y. Mean levels of BMI also increased, from 25.3 to 26.5. The prevalence of overweight also increased markedly for children and adolescents in the United States over the same period. Before these observations, previous surveys had shown only slight increases from the 1960s to 1980 for both children and adults.^{1–3}

Little is known about the reasons for these increases in obesity in the population, although similar trends have been seen elsewhere in the world as well. Within a given population, these increases are likely to be responses to changes in environmental and social factors, rather than to genetic changes in the popula-

tion or to the effects of immigration and emigration. As Rose has pointed out, increases in the prevalence of an elevated level of a characteristic such as blood pressure or BMI tend to be associated with shifts in the distribution of that characteristic in the entire population.^{4–6}

Several general models for the increase in the population prevalence of obesity are possible. One subgroup of the population may be heavier than the comparable subgroup in the past, with little change in the rest of the population. Or all subgroups of the population might be heavier than the comparable subgroups were previously. Because both these changes, as well as many other types of changes, would lead to increases both in the prevalence of overweight and in mean BMI, these increases do not in themselves show exactly what the underlying changes in the population distribution of BMI might be. The Rose model would suggest that increases in the prevalence of obesity are likely to be related to changes in the distribution of BMI in the population as a whole.

The objective of this paper is to apply graphical methods to investigate the changes in the distribution of BMI in the United States, using data from cross-sectional nationally representative health examination surveys. Because these comparisons are based on successive cross-sectional surveys, each of which examined a different sample of individuals, these

*Correspondence: KM Flegal, National Center for Health Statistics, 6525 Belcrest Rd., Room 900, Hyattsville, MD 20782, USA.
E-mail: kmf2@cdc.gov
Received 16 August 1999; revised 6 January 2000; accepted 10 January 2000



increases over time do not show changes on an individual level. However, the differences in the distribution of BMI between two surveys can be examined.

Because we found no standard approach in the literature for comparing distributions of BMI, we reviewed a number of graphical methods. We selected Tukey mean-difference (m-d plots)⁷ to compare the distribution of BMI in NHANES III to the distribution of BMI in prior surveys for the same sex-age categories. Since these plots may be unfamiliar, we first describe the construction and use of m-d plots and give some examples. We then apply this method to data for the US population. We previously published some examples of m-d plots for a few selected age-sex groups for adults from the first phase of NHANES III (1988-91) and for children and adolescents.^{3,8}

Methods

NHANES III

The third National Health and Nutrition Examination Survey (NHANES III) was conducted from 1988 to 1994 by the National Center for Health Statistics (NCHS) of the Centers for Disease Control and Prevention. A nationally representative sample of the US civilian non-institutionalized population was selected using a complex, stratified, multistage probability cluster sampling design. Informed consent was obtained from all respondents and the protocol was reviewed and approved by the NCHS NHANES Institutional Review Board. A home interview was followed by a physical examination in a mobile examination center. A description of the plan and operation of the survey has been published.⁹

NHANES III is the most recent in a series of cross-sectional health examination surveys carried out by NCHS. Although each survey provides representative national data, different individuals are examined each time. In all surveys, weight and height were measured using standardized techniques and equipment.¹ Body mass index (BMI) was calculated as weight (kg) divided by the square of height (m). For adults, obesity was defined as a BMI value ≥ 30.0 . For children and adolescents, overweight was defined as a BMI value equal to or greater than the preliminary month-of-age-and sex-specific 95th percentile values from the revision of the NCHS growth charts.

For adults aged 20-74 y, we compared the distribution of BMI in NHANES III with the distribution of BMI in the preceding survey, NHANES II, which was conducted from 1976 to 1980 using a similar design and similar methods. A description of that survey has also been published.¹⁰ In NHANES III there was no upper age limit. However, because for NHANES II, the upper age limit was 74 y, comparisons for adults

are limited to the age range 20-74 y. Women who were pregnant at the time of the examination are excluded from these analyses. Unweighted sample sizes by sex and age group for adults are presented in Table 1.

For children and adolescents aged 6-17 y, we compared the distribution of BMI in NHANES III with the distribution of BMI in two earlier surveys in the series. Cycles II and III of the National Health Examination Survey (NHES II, 1963-65; NHES III, 1966-70). Descriptions of these surveys have also been published.^{11,12} NHES II was limited to children aged 6-11 y and NHES III was limited to children and adolescents aged 12-17 y. Because the prevalence of overweight in these two earlier surveys was similar to that in NHANES II and because these surveys had large sample sizes for children and adolescents, these surveys were selected as the basis

Table 1 Sample sizes by sex, age and survey for adults aged 20-74 y

| Age-sex group | NHANES II | NHANES III |
|---------------|-----------|------------|
| <i>Men</i> | | |
| 20-29 y | 1261 | 1639 |
| 30-39 y | 871 | 1470 |
| 40-49 y | 695 | 1222 |
| 50-59 y | 691 | 858 |
| 60-69 y | 1579 | 1179 |
| 70-74 y | 507 | 548 |
| <i>Women</i> | | |
| 20-29 y | 1290 | 1669 |
| 30-39 y | 964 | 1776 |
| 40-49 y | 765 | 1358 |
| 50-59 y | 793 | 1005 |
| 60-69 y | 1729 | 1174 |
| 70-74 y | 620 | 570 |

Table 2 Sample sizes by age, sex and survey for children and adolescents aged 6-17 y

| Age-sex group | NHES II/III | NHANES III |
|----------------|-------------|------------|
| <i>Males</i> | | |
| 6 y | 575 | 282 |
| 7 y | 632 | 269 |
| 8 y | 618 | 266 |
| 9 y | 603 | 279 |
| 10 y | 576 | 297 |
| 11 y | 595 | 280 |
| 12 y | 643 | 203 |
| 13 y | 626 | 187 |
| 14 y | 618 | 188 |
| 15 y | 613 | 187 |
| 16 y | 556 | 194 |
| 17 y | 458 | 196 |
| <i>Females</i> | | |
| 6 y | 536 | 272 |
| 7 y | 609 | 274 |
| 8 y | 613 | 247 |
| 9 y | 581 | 280 |
| 10 y | 584 | 258 |
| 11 y | 525 | 275 |
| 12 y | 547 | 236 |
| 13 y | 582 | 220 |
| 14 y | 586 | 220 |
| 15 y | 503 | 197 |
| 16 y | 536 | 215 |
| 17 y | 442 | 202 |

for comparison with children and adolescents from NHANES III. Unweighted sample sizes for children and adolescents are presented in Table 2. Pregnant girls aged 17 ($n=15$) in NHANES III were excluded from these analyses: pregnancy status was not available for younger girls or for earlier surveys.

Mean-difference plots

To examine the changes in the distribution of BMI, we used Tukey mean-difference plots (m-d plots) to compare the distributions from the two surveys.⁷ A mean-difference plot is a graphical method for comparing distributions that is related to the quantile-quantile (q-q) plot of Wilk and Gnanadesikan.⁷ For both types of plots, shifts in distributions are investigated by comparing corresponding quantiles (for example, percentiles) from two distributions. To construct a q-q plot, the quantiles from one distribution are graphed against the corresponding quantiles from the other distribution. To construct an m-d plot, the differences between the corresponding quantiles are graphed on the y axis against the means of the same quantiles on the x axis. The points on the plot represent the differences between two distributions at a given quantile level plotted against the means of the same quantile level from the two distributions. For example, if the value of the 40th percentile for one distribution was 21.5 and for the other distribution was 20.0, the plot would include a point for that percentile that showed the difference between the two values (1.5) plotted against the mean of the two values (20.75). If the two distributions were the same, the quantiles would be equal and the differences would all be zero. The location of the plotted point on the y axis shows the direction and magnitude of the shift. The shift is assessed by judging differences from the horizontal line representing zero.

Another graphical method sometimes used to compare distributions is a plot that shows the cumulative percent of the population from each distribution at each BMI value.¹³⁻¹⁵ The information in a plot of cumulative distributions is similar to that contained in an m-d plot. However, the plot of cumulative distributions can be difficult to interpret and a clearer visual assessment of the type of shifts between the two distributions can often be derived from the m-d plot. The m-d plot shows the differences more explicitly than the cumulative distribution plot and is easier to assess visually because of the horizontal orientation. It also shows the numeric value of the shift more clearly.

Three simplified examples showing several types of possible differences between two distributions are presented in Figures 1, 2 and 3, along with the m-d plots and cumulative distribution plots corresponding to those differences. In the first example, both distributions have the same shape but the second distribution is shifted rightward (upward) by a constant amount relative to the first distribution (Figure 1). The constant shift appears in the m-d plot as a

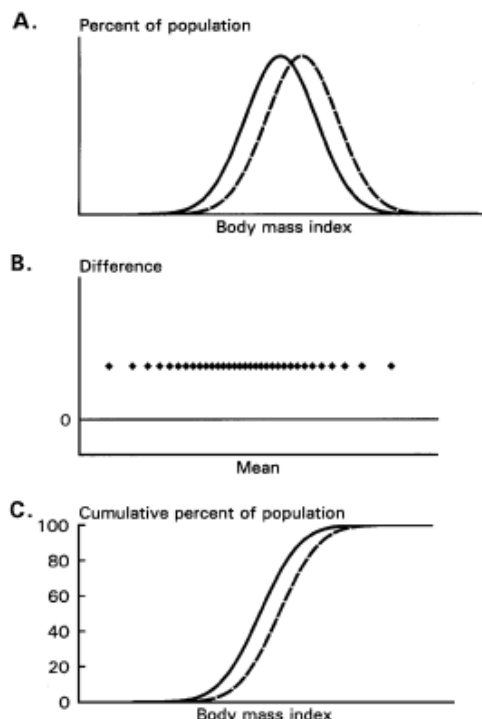


Figure 1 Schematic representation of a constant rightward (upward) shift between two distributions. (a) Frequency distributions, (b) Corresponding m-d plot, (c) corresponding cumulative distribution plot.

constant difference between percentiles of the two distributions. The amount of the constant shift can be read directly from the m-d plot axis.

The second example also shows two distributions (Figure 2). At the lower end, the two distributions are very similar and correspond closely, but at the upper end, one distribution is more skewed to the right than the other. In the corresponding m-d plot, the differences corresponding to the lower percentiles are close to zero, showing no shift of the distribution at the lower percentiles, and increase progressively at higher percentiles, showing the increasing skewness.

The third example (Figure 3) combines the features of the first and second examples. The second distribution, relative to the first, is both shifted upward and more skewed. The corresponding m-d plot shows both the shift at lower percentiles and an increasing difference at upper percentiles indicating increased skewness as well.

These three figures provide schematic representations of three possible changes in the distribution of BMI, all of which would result both in an increased mean BMI and in an increase in the prevalence of obesity. These are presented simply as examples to show how m-d plots represent some specific changes

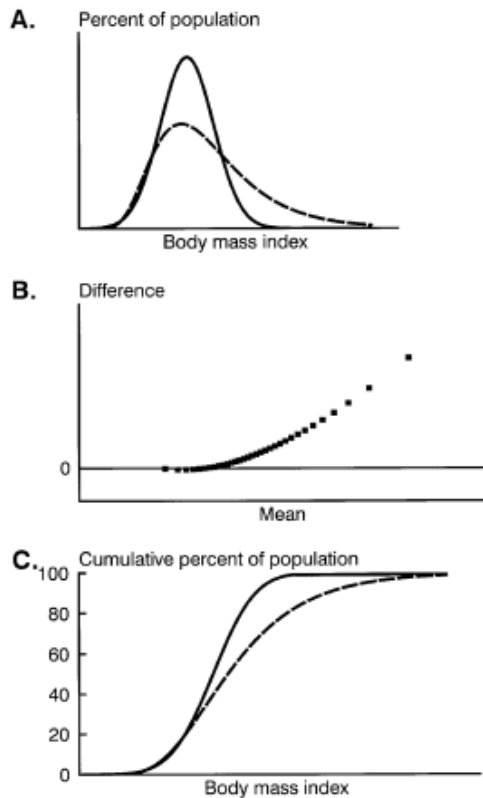


Figure 2 Schematic representation of a change to increased skewness. (a) Frequency distributions. (b) Corresponding m-d plot. (c) Corresponding cumulative distribution plot.

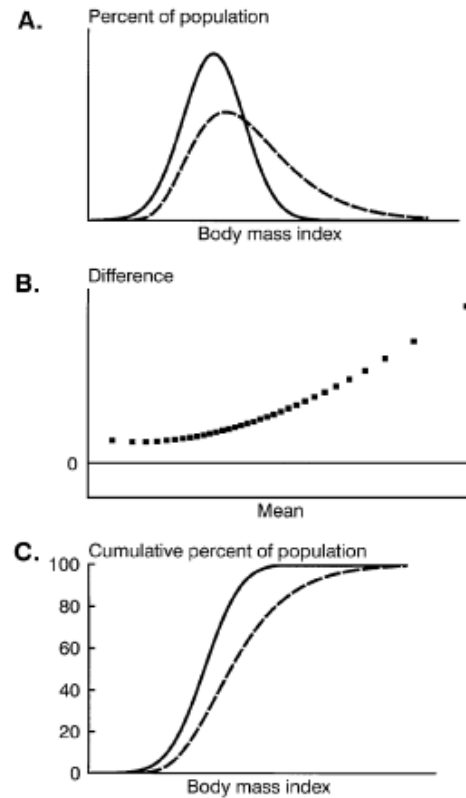


Figure 3 Schematic representation of both a rightward shift and increased skewness. (a) Frequency distributions. (b) Corresponding m-d plot. (c) Corresponding cumulative distribution plot.

between two distributions. Many other types of changes could also occur.

As a real-life example, a frequency distribution plot for men aged 40–49 y is presented in Figure 4 with the corresponding m–d plot and cumulative distribution plot. The frequency distribution is sensitive to the choice of intervals for the plots. In this example, the interval is 1 BMI unit. A wider interval would produce a smoother plot but lose detail and obscure some features of the distributions. A finer interval would produce a more detailed but more jagged plot. In contrast, the quantiles used for m–d plots are not affected by the choice of intervals. Although with smaller intervals between quantiles the number of points on the plot will increase, the value of any given point such as the 90th percentile is not affected by the choice of interval. As can be seen from this example and the preceding examples, in both the frequency distribution and the cumulative distribution plot, it is difficult to evaluate changes at the extremes of the distribution. It is also difficult to assess how constant or how variable the shift might be across the

distributions. The m–d plot permits evaluation of both these features.

Data analysis and statistical methods

For adults, age was defined as age in years at the time of the household interview, which generally preceded the examination by 2–3 weeks. The age groups used were 20–29 y, 30–39 y, 40–49 y, 50–59 y, 60–69 y and 70–74 y. For children and adolescents ages 6–17 y age was defined as age at the time of the examination. Single year of age groupings were used for children and adolescents because over this age range BMI increases with age and the BMI values for children of different ages are not directly comparable. In the age range 6–17 y, findings should be interpreted cautiously because the unweighted sample sizes for the NHANES III data are small, particularly for adolescents.

Statistical analyses were carried out using SAS and SUDAAN.^{16,17} For each survey, sampling weights

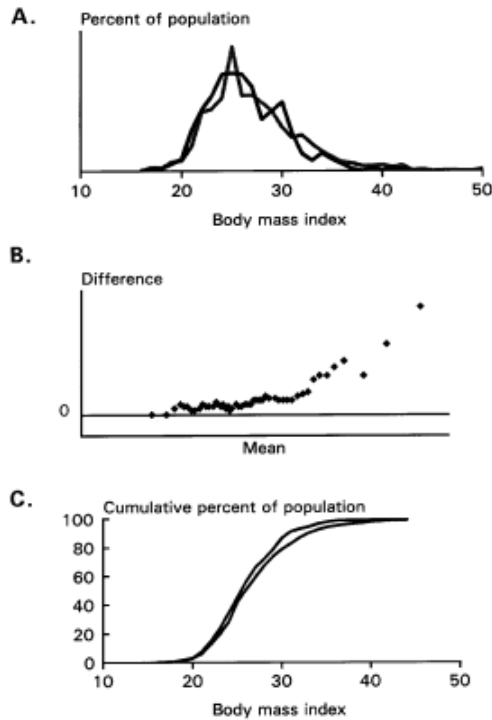


Figure 4 Example of frequency distributions from NHANES II and NHANES III for men aged 40–49 y. (a) Frequency distributions, (b) Corresponding m–d plot, (c) Corresponding cumulative distribution plot.

have been calculated that take into account the unequal probabilities of selection resulting from the sample design and from planned oversampling of certain subgroups. Confidence intervals were calcu-

lated using variance estimates that took into account the complex sample design of the surveys.

Weighted percentiles were calculated using the sampling weights. We calculated the even percentile values for each distribution (2nd, 4th, 6th percentile, etc., up to the 98th percentile). For each percentile level, we then calculated the mean of the corresponding percentile values from the two distributions and the difference between the corresponding percentile values from the two distributions, calculated as NHANES III minus the other survey. Each point on the plot represents the mean (on the x axis) and the difference (on the y axis) for an even percentile from the 2nd to the 98th percentile. Data were omitted from the m–d plots for one adolescent male age 14 y who had both a very high BMI (49.4) and a high sampling weight, making this single point unduly influential. For children and adolescents, analyses were repeated limiting the sample to white children and adolescents in NHES and to non-Hispanic white children and adolescents in NHANES III. The results were similar to those found including all race-ethnic groups (data not shown).

Results

Increases in mean BMI and the prevalence of obesity

For adults the increases in mean and median BMI and in the prevalence of obesity (BMI ≥ 30.0) between the two surveys are shown in Table 3. For children and adolescents, the differences in mean and median BMI are presented in Table 4 by single year of age, and the differences in the prevalence of overweight are presented in Table 5 for broader age groups. For every age–sex group, both mean and median BMI and the prevalence of overweight or obesity were greater in NHANES III than in the previous survey, suggesting that for both children and adults the distribution of

Table 3 Increases in mean and median BMI and in the prevalence of obesity (BMI ≥ 30.0) for adults aged 20–74 y. United States, 1976–80 to 1988–94

| Age–sex group | Mean BMI | | | Median BMI | | | Prevalence of obesity (BMI ≥ 30.0) | | |
|---------------|----------|------|--------------------|------------|------|--------------------|--|------|--------------------|
| | NHANES | | | NHANES | | | NHANES | | |
| | II | III | Increase* (95% CI) | II | III | Increase* (95% CI) | II | III | Increase* (95% CI) |
| Men | | | | | | | | | |
| 20–29 y | 24.3 | 25.2 | 0.9 (0.5–1.3) | 23.7 | 24.3 | 0.6 (0.2–1.1) | 8.1 | 12.5 | 4.4 (1.6–7.2) |
| 30–39 y | 25.6 | 26.5 | 0.9 (0.4–1.4) | 25.1 | 25.7 | 0.6 (0.2–1.0) | 11.9 | 17.2 | 5.3 (2.0–8.7) |
| 40–49 y | 26.4 | 27.3 | 1.0 (0.4–1.5) | 25.9 | 26.3 | 0.5 (–0.1–1.1) | 15.4 | 23.1 | 7.7 (3.3–12.2) |
| 50–59 y | 26.2 | 27.8 | 1.6 (1.2–2.1) | 25.8 | 27.2 | 1.3 (0.8–1.8) | 14.1 | 28.9 | 14.8 (10.4–19.1) |
| 60–69 y | 25.7 | 27.4 | 1.6 (1.2–2.1) | 25.7 | 27.1 | 1.5 (0.8–2.1) | 13.5 | 24.8 | 11.4 (6.5–16.3) |
| 70–74 y | 25.4 | 26.8 | 1.3 (0.6–2.0) | 25.2 | 26.1 | 0.9 (0.2–1.6) | 13.4 | 21.1 | 7.8 (3.3–14.3) |
| Women | | | | | | | | | |
| 20–29 y | 23.1 | 24.3 | 1.1 (0.6–1.7) | 22.0 | 22.7 | 0.7 (0.1–1.3) | 8.9 | 14.6 | 5.7 (2.1–9.3) |
| 30–39 y | 24.9 | 26.3 | 1.4 (0.7–2.2) | 23.1 | 24.4 | 1.2 (0.5–2.0) | 16.8 | 25.8 | 9.1 (4.5–13.6) |
| 40–49 y | 25.7 | 27.0 | 1.3 (0.5–2.1) | 24.1 | 25.4 | 1.3 (0.5–2.0) | 18.1 | 26.9 | 8.8 (3.9–13.7) |
| 50–59 y | 26.5 | 28.4 | 1.9 (1.2–2.6) | 25.3 | 27.1 | 1.9 (0.9–2.8) | 22.4 | 35.6 | 13.2 (8.0–18.3) |
| 60–69 y | 26.5 | 27.6 | 1.1 (0.6–1.6) | 25.6 | 26.6 | 1.0 (0.4–1.5) | 22.0 | 29.8 | 7.9 (3.9–11.8) |
| 70–74 y | 26.5 | 27.3 | 0.9 (0.2–1.5) | 25.9 | 26.3 | 0.3 (–0.4–1.1) | 18.4 | 25.7 | 7.2 (3.1–11.4) |

*Numbers may not add up exactly due to rounding.



Table 4 Increases in mean and median BMI for children and adolescents aged 6–17 y. United States, 1963–70 to 1988–94

| Age–sex | Mean BMI | | | Median BMI | | |
|----------------|-------------|------------|--------------------|-------------|------------|--------------------|
| | NHES II/III | NHANES III | Increase* (95% CI) | NHES II/III | NHANES III | Increase* (95% CI) |
| <i>Males</i> | | | | | | |
| 6 y | 15.6 | 16.3 | 0.7 (0.2–1.3) | 15.4 | 15.7 | 0.3 (–0.2–0.7) |
| 7 y | 15.9 | 16.5 | 0.6 (0.2–1.0) | 15.6 | 15.8 | 0.2 (–0.2–0.6) |
| 8 y | 16.3 | 17.3 | 1.0 (0.3–1.7) | 15.9 | 16.3 | 0.4 (–0.2–1.1) |
| 9 y | 16.9 | 18.0 | 1.1 (0.3–2.0) | 16.2 | 17.0 | 0.8 (–0.1–1.7) |
| 10 y | 17.1 | 18.4 | 1.3 (0.7–1.8) | 16.5 | 17.7 | 1.2 (0.4–1.9) |
| 11 y | 17.9 | 19.4 | 1.4 (0.8–2.0) | 17.2 | 18.3 | 1.1 (0.3–1.8) |
| 12 y | 18.4 | 20.1 | 1.7 (1.2–2.3) | 17.7 | 19.0 | 1.3 (0.6–2.0) |
| 13 y | 19.4 | 20.5 | 1.1 (0.5–1.8) | 18.7 | 19.5 | 0.7 (–0.1–1.5) |
| 14 y | 20.2 | 22.3 | 2.1 (–0.2–4.3) | 19.6 | 20.8 | 1.2 (–0.1–2.4) |
| 15 y | 20.9 | 22.3 | 1.4 (0.4–2.4) | 20.3 | 20.9 | 0.6 (–0.6–1.8) |
| 16 y | 21.3 | 22.3 | 1.0 (0.1–2.0) | 20.7 | 21.1 | 0.4 (–0.6–1.5) |
| 17 y | 22.1 | 23.4 | 1.4 (0.5–2.2) | 21.4 | 21.8 | 0.4 (–0.6–1.4) |
| <i>Females</i> | | | | | | |
| 6 y | 15.4 | 16.1 | 0.6 (0.1–1.2) | 15.2 | 15.3 | 0.1 (–0.3–0.6) |
| 7 y | 15.8 | 16.9 | 1.1 (0.5–1.7) | 15.5 | 15.8 | 0.4 (–0.2–1.0) |
| 8 y | 16.4 | 17.3 | 0.9 (0.3–1.5) | 16.0 | 16.6 | 0.6 (–0.2–1.4) |
| 9 y | 17.0 | 18.2 | 1.2 (0.3–2.2) | 16.3 | 16.6 | 0.3 (–0.5–1.1) |
| 10 y | 17.6 | 18.4 | 0.8 (–0.1–1.8) | 16.9 | 17.3 | 0.4 (–0.3–1.2) |
| 11 y | 18.2 | 19.4 | 1.2 (0.4–2.0) | 17.5 | 18.5 | 1.0 (–0.3–2.2) |
| 12 y | 19.2 | 20.2 | 1.0 (–0.0–1.9) | 18.5 | 19.4 | 0.9 (–0.4–2.2) |
| 13 y | 19.9 | 21.8 | 1.9 (0.7–3.1) | 19.3 | 20.5 | 1.2 (–0.2–2.5) |
| 14 y | 20.8 | 22.4 | 1.7 (0.7–2.7) | 20.2 | 21.1 | 0.9 (–0.5–2.3) |
| 15 y | 21.4 | 22.0 | 0.5 (–0.4–1.4) | 20.6 | 21.1 | 0.4 (–0.2–1.1) |
| 16 y | 21.9 | 23.2 | 1.2 (0.3–2.2) | 21.0 | 21.2 | 0.2 (–0.4–0.8) |
| 17 y | 21.7 | 23.3 | 1.6 (0.6–2.6) | 21.0 | 22.2 | 1.2 (0.1–2.3) |

*Numbers may not add up exactly due to rounding.

Table 5 Mean prevalence of overweight (BMI ≥ 95th percentile) for children and adolescents aged 6–17 y. United States, 1963–70 to 1988–94

| Age–sex group | Prevalence (%) of overweight* | | |
|----------------|-------------------------------|------------|-------------------|
| | NHES II/III | NHANES III | Increase (95% CI) |
| <i>Males</i> | | | |
| 6–8 y | 3.4 | 10.8 | 7.3 (3.5–11.1) |
| 9–11 y | 4.6 | 12.8 | 8.2 (4.3–12.0) |
| 12–14 y | 4.7 | 11.9 | 7.2 (2.5–11.9) |
| 15–17 y | 4.2 | 12.0 | 7.8 (3.4–12.2) |
| <i>Females</i> | | | |
| 6–8 y | 4.1 | 11.0 | 6.9 (3.0–10.9) |
| 9–11 y | 4.9 | 11.0 | 6.1 (2.4–9.7) |
| 12–14 y | 4.4 | 11.7 | 7.3 (2.2–12.4) |
| 15–17 y | 5.1 | 8.7 | 3.6 (1.2–6.1) |

*Based on sex and month-of-age-specific percentile cutoffs from revised US growth charts (preliminary data).

BMI values in NHANES III differs from previous surveys. For the changes in mean BMI and the changes in the prevalence of overweight, 95% confidence intervals generally did not include zero. For children and adolescents, the median BMIs were always greater in NHANES III than in NHES II/III although the confidence intervals generally included zero.

Mean–difference plots

M–d plots comparing the distribution of BMI in NHANES III with the distribution of BMI in NHES

II are shown in Figure 5 by single year of age for children aged 6–11 y. The y-axis shows the value of the difference between corresponding percentiles of the two distributions and can be interpreted as showing the direction and distance of any shift between the distributions. The shape of the line formed by the points suggests the nature of the shift. These are qualitative graphical methods designed to summarize and display differences between two distributions. Because of this and because of the small sample sizes for some age groups the general patterns are more important than the individual details.

For 6-y-old children, both male and female, the m–d plots show that the distributions of BMI from the two surveys are virtually identical over most of the range. Most of the differences are very close to zero. However, at the upper (higher) end of the distribution, the differences begin to increase progressively until at the highest percentiles, the percentiles for NHANES III are 4–6 BMI units greater than the corresponding NHES II percentiles. These large differences show a marked upward shift and increased skewness at the upper end of the distribution, while most of the distribution is unchanged.

Although there are differences in detail, the patterns for children aged 6–11 y are broadly similar. Most of the plots suggest that the distributions of BMI in the two surveys have changed little if at all at the lower end, where the mean differences are close to and sometimes less than zero. However, at the upper end of the distribution all plots, except for 11-year-old

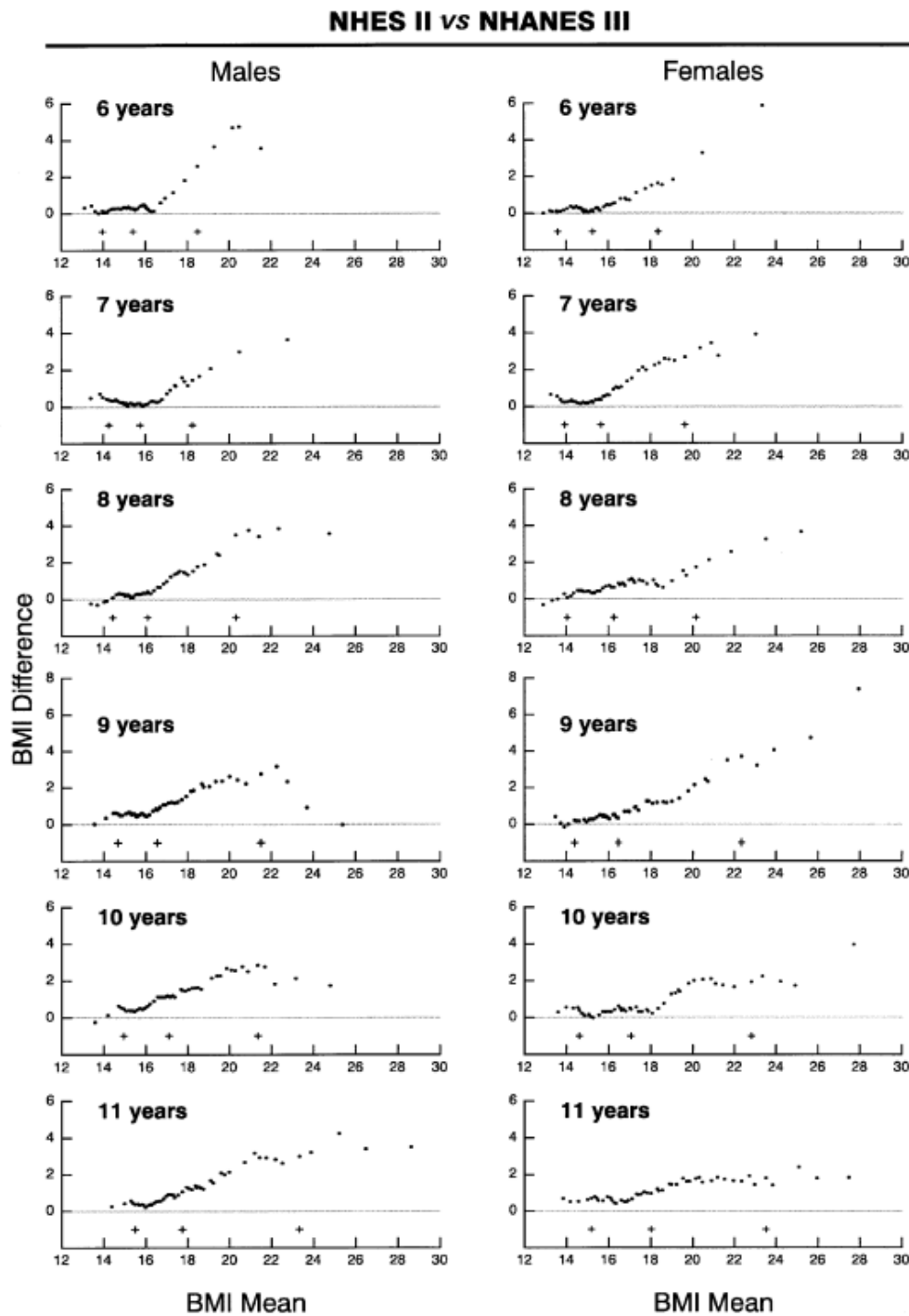


Figure 5 Mean-difference plots for the distribution of body mass index from NHES II and NHANES III, for children aged 6–11 y, by sex and age group. The 10th, 50th and 90th percentiles are marked by a + sign.



NHES III vs NHANES III

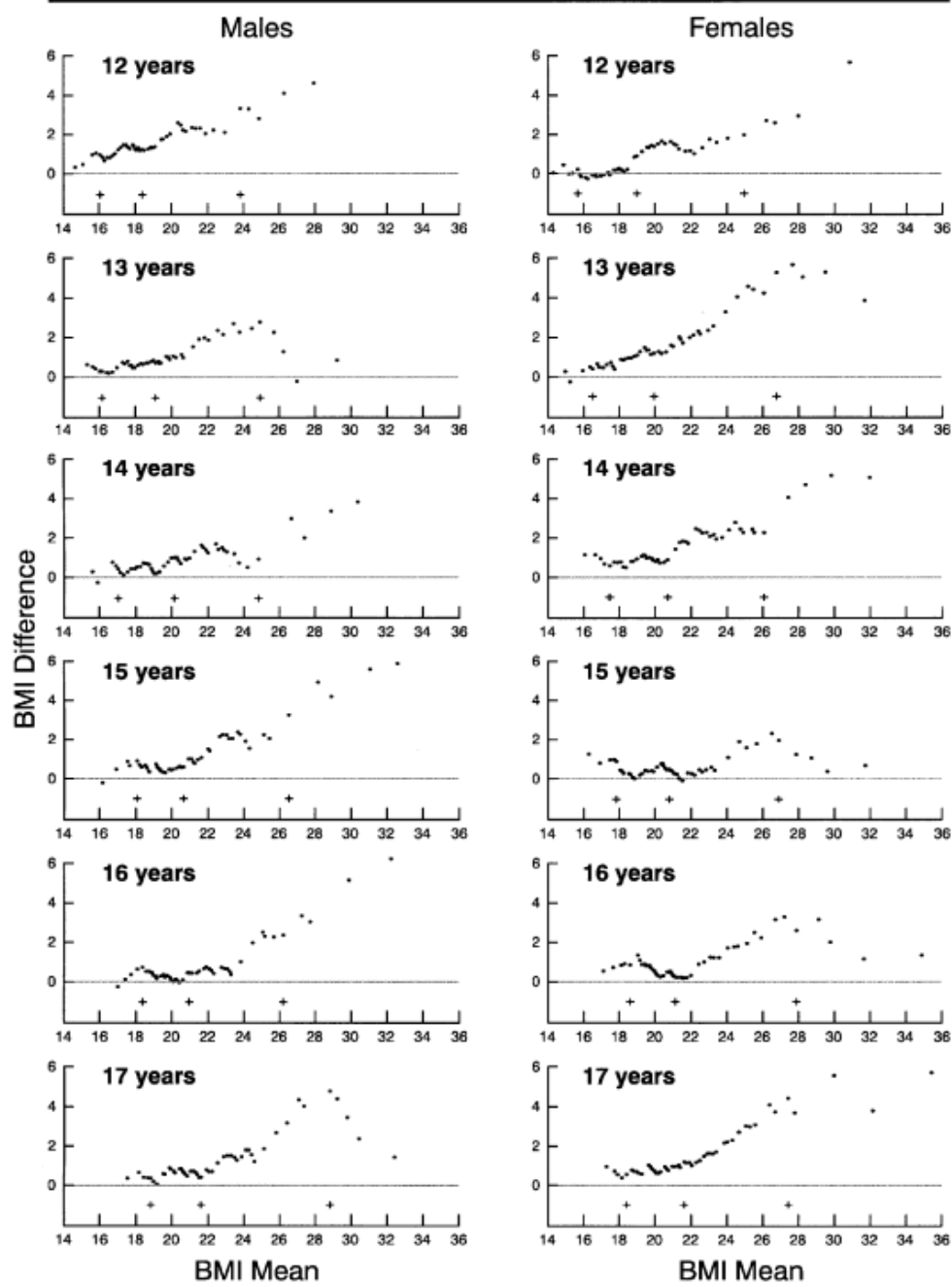


Figure 6 Mean-difference plots for the distribution of body mass index from NHES III and NHANES III, for adolescents aged 12–17 y, by sex and age group. The 10th, 50th and 90th percentiles are marked by a + sign.

girls, show a large difference of at least 2 BMI units and often much more. This shows pronounced increases in skewness with large shifts at the upper end of the distribution. Thus the changes in the distribution of BMI for children between NHES II and NHANES III are characterized by little or no difference at the lower end of the distribution, increasing skewness and large changes at the upper end of the distribution.

M-d plots comparing the distribution of BMI in NHANES III with the distribution of BMI in NHES III are shown in Figure 6 by single year of age for adolescents aged 12–17 y. Here again, the pattern is similar with little change at the lowest part of the distribution but strikingly large changes of 2–6 BMI units at the upper percentiles for most groups, except for girls aged 15 and 16 y.

M-d plots comparing the distribution of BMI in NHANES III with the distribution of BMI in NHANES II are shown in Figure 7 for men and women aged 20–74 y by 10 year age groups. Although the details differ between sex–age groups, there are some broad similarities. In almost every case, the percentiles from NHANES III are greater than the corresponding percentiles from NHANES II. This shows that there is some upward shift in the distribution of BMI for every sex–age group. All age–sex groups also show greater differences in the upper portion of the distribution. This shows that there is also increased skewness in the distribution of BMI for every sex–age group. The magnitude of the differences varies from group to group. However, in most groups, the largest differences are approximately 3–4 BMI units and are seen at the highest percentiles of the distribution.

For younger men (aged 20–49 y), the differences at the lower end of the distribution are small (< 1 BMI unit) and fairly constant. However, at the upper end of the distribution, the differences are much larger and are progressively greater at higher percentile levels. This pattern shows that, at these ages, there is a slight upward shift of the entire distribution and in addition a large increase at the upper end of the distribution, so that the distribution is also becoming more skewed.

For men in the older age ranges (50–74 y), the pattern is similar, but the degree of the shift is greater at lower levels and less at higher percentiles than for the younger age groups. This pattern suggests that for older men, the upward shift is stronger and the increased skewness at the upper end of the distribution is less than for younger men.

For women, the pattern is different. For most age groups of women, the differences increase steadily across percentiles. This shows that the entire distribution is shifting upward and becoming progressively more skewed at higher levels. Compared with men in the same age range, the shifts in the distribution for women are greater at lower percentiles for the age range 20–49 y.

Discussion

Increases in mean and median BMI and in the prevalence of BMI ≥ 30.0 all show that the distribution of BMI for adults in NHANES III differs in some way from the distribution of BMI in NHANES II. Similar findings for mean and median BMI and the prevalence of BMI ≥ 95 th percentile also show that the distribution of BMI for children and adolescents in NHANES III differs from the distribution of BMI for the corresponding age–sex groups in NHES II and III. However, these findings do not in themselves give a complete picture of how the distributions of BMI have changed over time.

The changes in the distributions can be represented graphically by mean–difference plots. These graphical methods by their nature include a very large number of individual data points, as compared with summary statistics such as the mean or standard deviation. The purpose of these methods is not to focus closely on any given individual data point, but rather to depict graphically the overall differences in the two distributions. These methods are not suited, for instance, clearly to identify with certainty a particular value of BMI above which the distribution shifts. However, they can show broad qualitative differences between the distributions.

These plots show that for both adults and children, the distribution of BMI is becoming more skewed in all sex–age groups. The heaviest subgroup of the population is much heavier in NHANES III than was the heaviest subgroup in previous surveys. In addition, for adults, the entire distribution of BMI is shifted upward in NHANES III relative to NHANES II. However, these plots show that this is not a uniform shift across the whole distribution.

The degree of these changes in the distribution differs by age. For younger children and to some extent for adolescents, the lower portion of the distribution seems to be anchored with no substantive change over time. In these age groups, the principal effect is increased positive skewness at the upper part of the distribution. For younger men (aged 20–49 y) the principal change is also increased positive skewness at the upper part of the distribution, but there is in addition a slight upward shift that tends to be small and fairly consistent. The differences in the lower portion of the distribution are small (< 1 BMI unit). For older men (aged 50–74 y), however, the upward shift is more pronounced and the shift at the upper end, indicating skewness, less pronounced. For women, there is a more gradual increase across the entire distribution, suggesting both an upward shift and increasing positive skewness.

Taken together, these observations suggest that for adults some factors causing increases in BMI are affecting the entire distribution of BMI, although the changes in the distribution of BMI are most marked at the upper end of the distributions. These plots show

NHANES II vs NHANES III

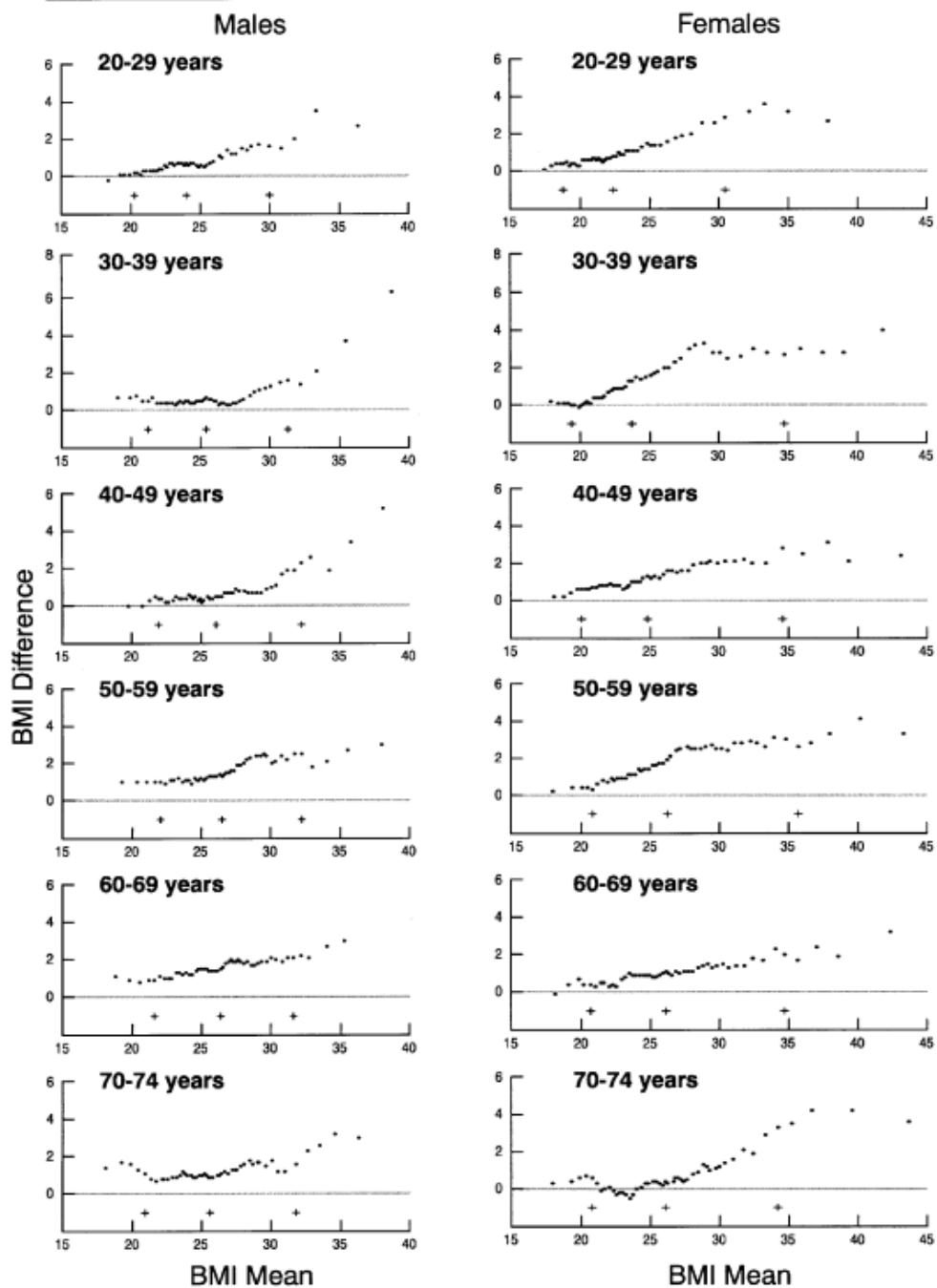


Figure 7 Mean-difference plots for the distribution of body mass index from NHANES II and NHANES III, for adults aged 20-74 y, by sex and age group. The 10th, 50th and 90th percentiles are marked by a + sign.

that for adults in all age-sex groups, almost the entire distribution of BMI appears to have shifted upward to some extent in NHANES III compared with the corresponding distributions in NHANES II. Thus these findings do not suggest that the increases in mean BMI and in the prevalence of BMI ≥ 30.0 in NHANES III, relative to NHANES II, are due solely to the heaviest people in NHANES III being heavier than the heaviest people in NHANES II. For adults, there is no part of the distribution that is clearly unresponsive to the influences that are causing the population increase in BMI and the prevalence of obesity. For older men there is almost a complete shift of the distribution.

These findings suggest that the causes for the increase in obesity should be sought in part at the population level rather than focusing on individuals. From the perspective of individual behaviors, changes in physical activity, energy intake and smoking initiation or cessation may lead to changes in energy balance and thus to the increases in overweight and obesity. However, these changes in behaviors are likely to be related to social and environmental changes that may affect the whole population. According to Rose, to find the determinants of prevalence and incidence rates, we need to study characteristics of populations, not of individuals.⁴ The social and environmental causes of changes in BMI are unclear but might include increased availability of food, changes in food composition, changes in patterns of food intake, fewer opportunities for physical activity, increased use of labor-saving devices, changes in cigarette smoking, and perhaps also changes in cultural and social attitudes and values that might directly or indirectly affect body weight.¹⁸⁻²⁰

The changes observed in the US affect the population broadly enough that they appear unlikely to be due to the effects on the population of immigration and emigration. Nor are they likely to be due to trends in the race-ethnic composition of the US population. The increases in the prevalence of overweight are as large or larger for non-Hispanic white children and adults than for other subgroups.¹⁻³ When the analyses were repeated limiting the sample to white children and adults, the general findings were unchanged. These changes might represent either a differential response to changing environmental conditions, or a subset increasingly exposed to changing environmental conditions, or possibly some combination of differential exposure and differential response. These data do not permit these effects to be disentangled. However, these findings are not inconsistent with a gene-environment interaction effect. The factors that are causing the increase in obesity appear to operate broadly and have an effect on the distribution of BMI for almost the entire adult population, but to have a stronger effect on the upper portion of the distribution, suggesting there may be a subgroup of the population that is more susceptible to these influences. These data also broadly support Rose's suggestion that changes in

the upper portion of the distribution reflect changes occurring in the whole distribution.⁵

This interpretation is less true for children, however. For children, particularly younger children, and to some extent also for adolescents and young men, the main effect appears to be on the upper part of the distribution of BMI. This suggests that, for these groups, the increase in prevalence is chiefly due to the heaviest individuals being heavier in NHANES III than were the heaviest individuals in earlier surveys. This is consistent with Price's findings for secular trends in overweight among young Danish inductees, in which BMI could be modeled as a mixture of distributions of a 'normal' and 'overweight' component.¹⁴ Secular trends in overweight among young Danish men could be explained by an increase in the proportion in the overweight component, with no change in the mean of either the normal or the overweight component, leading to the speculation that the increases in overweight might represent a differential response to environmental change, possibly mediated by genotype. Although methodological differences preclude exact comparisons, our findings for children, adolescents and young men, showing only small or no shifts in the lower part of the distribution but with a much larger shift in the upper portion of the distribution, are consistent with the notion of a differential response in these age groups. For older men, however, our data suggest a more uniform response affecting all portions of the distribution.

Many investigations of BMI phenotypes suggest that BMI can often be modeled as a mixture of several components, although findings about the number and characteristics of these components vary from study to study.^{14,21,22} These shifts over time in the distributions of BMI, which appear to vary by sex and age, may affect such phenotypic comparisons of BMI and could make it more difficult to find stable models for these distributions.^{23,24}

Rose and Day examined cross-sectional data on blood pressure, BMI, alcohol intake and sodium intake from 52 populations in the Intersalt study and found that the distributions of most health related characteristics appeared to move up and down as a whole so that the population mean predicted the number of deviant individuals.⁵ In the case of BMI, however, the distributions of BMI in the populations with higher values were not only shifted upward but also tended to become more skewed.⁶ This increasing skewness is similar to that seen in the NHANES data for adults, where small increases in mean and median BMI values are accompanied by greater than expected increases in the prevalence of high values because of the increasing skewness.

Within a given population, differences in characteristics such as serum cholesterol probably reflect principally genetic variability with some effects of behavioral differences. However, the large differences seen between populations in characteristics such as serum cholesterol suggest profound effects of social



and environmental determinants of such characteristics. The environment includes not just the physical environment, but also social and cultural forces affecting work, leisure, time allocation, food patterns and availability, psychological factors and many other determinants. These trends in BMI suggest the combination of both profound environmental determinants at work and a population with a high degree of susceptibility. Following the line of thought put forth by Rose,⁴⁻⁶ this would imply that the reasons for the increasing prevalence of obesity should be sought not by comparing obese and non-obese individuals but rather by seeking to understand what factors are causing increases in the population as a whole. Interventions may need to focus on the population rather than solely on the heaviest individuals.

References

- Kuczmarski RJ, Flegal KM, Campbell SM, Johnson CL. Increasing prevalence of overweight among US adults. The National Health and Nutrition Examination Surveys, 1960 to 1991. *JAMA* 1994; **272**: 205-211.
- Flegal KM, Carroll MD, Kuczmarski RJ, Johnson CL. Overweight and obesity in the United States: prevalence and trends, 1960-1994. *Int J Obes* 1998; **22**: 39-47.
- Troiano RP, Flegal KM. Overweight children and adolescents: description, epidemiology, and demographics. *Pediatrics* 1998; **101**: 497-504.
- Rose G. Sick individuals and sick populations. *Int J Epidemiol* 1985; **14**: 32-38.
- Rose G, Day S. The population mean predicts the number of deviant individuals. *Br Med J* 1990; **301**: 1031-1034.
- Rose G. Ancel Keys Lecture. *Circulation* 1991; **84**: 1405-1409.
- Cleveland WS. *Visualizing Data*. Hobart Press, Summit, NJ, 1993.
- Flegal KM. Trends in body weight and overweight in the U.S. population. *Nutr Rev* 1996; **54**: S97-S100.
- National Center for Health Statistics. Plan and operation of the Third National Health and Nutrition Examination Survey, 1988-1994. *Vital Health Stat* 1994; **32**: 1-405.
- National Center for Health Statistics, McDowell A, Engel A, Massey JT, Maurer K. Plan and Operation of the Second National Health and Nutrition Examination Survey 1976-80. *Vital Health Stat* 1981; **15**: 1-144.
- National Center for Health Statistics. Plan and operation of a health examination survey of U.S. youths 12-17 years of age. *Vital Health Stat* 1969; **1**: 1-80.
- National Center for Health Statistics. Plan, operation, and response results of a program of children's examinations. *Vital Health Stat* 1967; **1**: 1-56.
- Hodge AM, Dowse GK, Toelupe P, Collins VR, Imo T, Zimmet PZ. Dramatic increase in the prevalence of obesity in western Samoa over the 13 year period 1978-1991. *Int J Obes* 1994; **18**: 419-428.
- Price RA, Ness R, Sorensen TI. Changes in commingled body mass index distributions associated with secular trends in overweight among Danish young men. *Am J Epidemiol* 1991; **133**: 501-510.
- Price RA, Lunetta K, Ness R, Charles MA, Saad MF, Ravussin E, Bennett PH, Pettitt DJ, Knowler WC. Obesity in Pima Indians. Distribution characteristics and possible thresholds for genetic studies. *Int J Obes* 1992; **16**: 851-857.
- SAS Institute Inc. *SAS Procedures Guide, Version 6, 3rd edn*. SAS Institute Inc. Cary, NC: 1990.
- Shah BV, Barnwell BG, Bieler GS. *SUDAAN User's Manual, Release 6.40*. Research Triangle Institute: Research Triangle Park, NC, 1995.
- James WPT. A public health approach to the problem of obesity. *Int J Obes* 1995; **19**(Suppl 3): S37-S45.
- Prentice AM, Jebb SA. Obesity in Britain: gluttony or sloth? *Br Med J* 1995; **311**: 437-439.
- Flegal KM, Troiano RP, Pamuk ER, Kuczmarski RJ, Campbell SM. The influence of smoking cessation on the prevalence of overweight in the United States. *N Engl J Med* 1995; **333**: 1165-1170.
- Mitchell LE, Nirmala A, Rice T, Reddy PC, Rao DC. Commingling analysis of adiposity in an Indian population. *Int J Obes* 1994; **18**: 1-8.
- Moll PP, Bums TL, Lauer RM. The genetic and environmental sources of body mass index variability: the Muscatine Ponderosity Family Study. *Am J Hum Genet* 1991; **49**: 1243-1255.
- Price RA, Charles MA, Pettitt DJ, Knowler WC. Obesity in Pima Indians: genetic segregation analyses of body mass index complicated by temporal increases in obesity. *Hum Biol* 1994; **66**: 251-274.
- Price RA. Within birth cohort segregation analyses support recessive inheritance of body mass index in white and African-American families. *Int J Obes* 1996; **20**: 1044-1047.